

UNIVERSIDADE FEDERAL DE MINAS GERAIS

**EXPRESSÃO, CRISTALIZAÇÃO E DETERMINAÇÃO
ESTRUTURAL DE MACROMOLÉCULAS BIOLÓGICAS POR
DIFRAÇÃO DE RAIOS X POR MONOCRISTAL**

Ludmila Maria Diniz Leroy

Belo Horizonte

2017

Dados Internacionais de Catalogação na Publicação (CIP)

Leroy, Ludmila Maria Diniz.

L619e Expressão, cristalização e determinação estrutural de macromo-
léculas biológicas por difração de raios X por monocristal / Ludmila
Maria Diniz Leroy . – 2017.
118f. il.

Orientador: Carlos Basílio Pinheiro.
Coorientador: Ronaldo Alves Pinto Nagem
Dissertação (mestrado) – Universidade Federal de
Minas Gerais – Departamento de Física.
Bibliografia: f. 110 -118.

1. Cristalografia. 2. Difração de raios X. 3. Cristalografia de
superfícies. 4. Cristalografia de proteínas. I. Título. II. Pinheiro ,
Carlos Basílio. Título III. Nagem, Ronaldo Alves Pinto

CDU –548:53 (043)

Elaborada pela Biblioteca Professor Manoel Lopes de Siqueira da UFMG.



Universidade Federal de Minas Gerais

Instituto de Ciências Exatas - ICEx

Departamento de Física

EXPRESSÃO, CRISTALIZAÇÃO E DETERMINAÇÃO
ESTRUTURAL DE MACROMOLÉCULAS BIOLÓGICAS POR
DIFRAÇÃO DE RAIOS X POR MONOCRISTAL

Ludmila Maria Diniz Leroy

Dr. Carlos B. Pinheiro (Orientador)

Dr. Ronaldo A. P. Nagem (Coorientador)

Belo Horizonte

2017

“

**“AND IF I CLAIM TO BE A WISE MAN,
IT SURELY MEANS THAT I DON'T KNOW”**

- Kerry Livgren, Kansas

Agradecimentos

Agradecimentos formais são devidos às agências de fomento: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); e ao Laboratório Nacional de Luz Sincrotron pelo acesso à linha de luz MX2 e ao Laboratório Automatizado de Cristalização de Macromoléculas (RoboLab).

Ternamente, agradeço ao professor Dr. Carlos Basílio, pelos muitos anos de orientação, e a todos os meus colegas do LabCri, pelos ensinamentos, companheirismo e cafés.

Ao professor Dr. Ronaldo Nagem, pelo apoio como coorientador, e a todo o grupo do laboratório BioEst, especialmente à Dra. Simara Semíramis, pelos ensinamentos e pela paciência com meus desjeitos, e ao Dr. Samuel Guimarães, pelos conselhos.

De forma muito especial, agradeço a minha família pela inspiração e suporte, especialmente à minha mamis, Lucinéa, meu pai, Riviene, e ao meu irmão Rodrigo. Também aos meus queridos amigos, principalmente meus colegas da física. Reclamamos, brigamos, xingamos, mas rimos e bebemos, então, ‘teve bão’.

Por fim, agradeço à música, aos meus estudos para contrabaixo, ao meu professor, Hudson Cunha, e meus colegas do COINJ. As sextas feiras foram dias mais bonitos.

Resumo

A função e a forma de proteínas estão intimamente ligadas. A compreensão da estrutura tridimensional delas se faz então necessária na elucidação detalhada dos processos biológicos nos quais elas participam. A cristalografia é uma poderosa técnica empregada na determinação estrutural de proteínas. Vasto conhecimento de física e bioquímica são necessários na capacitação de um profissional apto a trabalhar em todas as etapas do processo envolvido na determinação estrutural por cristalografia de proteínas. O objetivo do presente trabalho foi visitar todas essas etapas, que incluem expressão proteica por técnicas de DNA recombinante, a purificação e a cristalização da proteína, experimentos de difração de raios X por monocristal (XRD), processamento de dados, solução da estrutura e análises do modelo proposto. Para realizar todas as etapas, três proteínas foram estudadas: uma esfingomielinase D (SMase D), o domínio E2 do papiloma vírus bovino tipo 1 (E2 de BPV-1) e o complexo NahK/NahL. O trabalho com cada uma das proteínas cobriu partes distintas do processo de determinação estrutural. Foi feita a expressão, purificação, cristalização e experimentos de XRD com cristais da SMase D; a estrutura do domínio E2 de BPV-1 foi elucidada a partir de dados de XRD e o modelo foi devidamente analisado; e os dados de XRD de um cristal de NahK/NahL, que indicaram uma desordem no mapa de densidade eletrônica experimental, foram devidamente reprocessados e analisados, resultando na proposta de um novo modelo para a estrutura do complexo.

Palavras chave: **Cristalografia de proteínas, difração de raios X, SMase D, E2 de BPV-1, complexo NahK/NahL.**

Abstract

Proteins' shape and function are intimately related. Hence, the comprehension of their tridimensional structures is necessary for the detailed elucidation of the biological processes in which they participate. Crystallography is a powerful technique applied on protein structure determination. Great knowledge of Physics and Biochemistry is needed when capacitating a professional to be able to work on every stage of the process of structure determination by protein crystallography. The present work aimed at visiting every one of those stages, including recombinant protein expression, protein purification and crystallization, X-ray diffraction experiments (XRD), data processing, structure solution and analyses. To cover the whole process, three proteins were studied: a sphingomyelinase D (SMase D), the bovine papillomavirus-1 E2 DNA binding domain (BPV-1 E2) and the NahK/NahL complex. Different parts of the process of structure determination were covered working with each of the proteins. SMase D was expressed, purified and crystallized and XRD experiments were performed with its crystals; the structure of BPV-1 E2 was elucidated from XRD data and the proposed model was properly analyzed; XRD data from a crystal of the NahK/NahL complex, which indicated an unordered region on the experimental electron density map, were reprocessed and analyzed, resulting on the proposition of a new structural model of the complex.

Keywords: Protein crystallography, X-ray diffraction, SMase D, BPV-1 E2, NahK/NahL complex.

Sumário

Capítulo 1

Introdução	1
1.1 Motivação.....	1
1.2 Objetivos	4

Capítulo 2

Materiais e Métodos	5
2.1 Proteínas	5
2.2 Produção e purificação de proteínas e análises.....	9
2.3 Cristalização	18
2.4 Teoria de espalhamento de raios X	21
2.5 Experimento de difração de raios X por monocristal.....	37
2.6 Solução e refinamento da estrutura.....	40

Capítulo 3

Esfingomielinase D	49
3.1 Introdução	49
3.2 Procedimentos e discussão	51
3.3 Conclusão.....	63

Capítulo 4

Complexo NahK/NahL.....	65
4.1 Introdução.....	65
4.2 Procedimentos e discussão.....	69
4.3 Conclusão.....	79
 Capítulo 5	
BPV E2.....	81
5.1 Introdução.....	81
5.2 Procedimentos e discussão.....	82
5.3 Conclusão.....	92
 Capítulo 6	
Conclusão e perspectivas.....	94
Apêndices.....	96
Referências.....	102

Introdução

1.1 Motivação

Proteínas são as macromoléculas biológicas que constituem o maquinário metabólico responsável pela manutenção da vida. Inúmeras funções são cumpridas por essas moléculas, desde as mais simples, como composição de estruturas celulares [1], até as mais complexas, como a produção de novas proteínas no ribossomo [2]. A compreensão desses mecanismos são de grande interesse científico e, portanto, alvo de inúmeros estudos envolvendo áreas da ciência da vida, medicina, química, física, engenharias e ciência da computação.

A estrutura tridimensional das proteínas está intimamente ligada à sua função de tal forma que sua determinação é essencial para a elucidação dos processos biológicos fundamentais. A determinação estrutural é, muitas vezes, um processo trabalhoso no qual várias técnicas experimentais podem ser utilizadas, tais como a cristalografia de proteínas, a ressonância magnética nuclear e, mais recentemente, a microscopia eletrônica. Notoriamente, a cristalografia de proteínas se mostra a mais poderosa ferramenta na investigação dessas estruturas, sendo hoje responsável por mais de 100 mil estruturas resolvidas depositadas no banco de dados, o *Protein Data Bank* (PDB) [3]. Essa técnica requer que a proteína de interesse seja cristalizada e que os cristais obtidos sejam adequados para experimentos de difração de raios X.

A obtenção de proteínas para cristalização é um processo longo que envolve várias etapas, que incluem: (i) isolar o gene que guarda informação da sequência peptídica desejada; (ii) clonar o gene em um vetor de expressão; (iii) inserir o

vetor em um organismo (*e. g. Escherichia coli*); (iv) fazer a cultura do organismo que recebeu o vetor; (v) induzir a expressão proteica; (vi) fazer a lise celular; e, finalmente, (vii) purificar a proteína de interesse. Algumas estratégias experimentais são empregadas na obtenção da proteína de interesse em quantidade suficiente para ensaios de cristalização. Vasto conhecimento de bioquímica e biologia molecular, além do domínio das diversas técnicas utilizadas na produção de uma proteína recombinante são necessários para a completa execução do processo.

Ainda que a obtenção de proteínas seja um processo complexo e demorado, é a cristalização o grande gargalo para a realização de estudos estruturais usando cristalografia. Os protocolos de cristalização utilizados atualmente consistem em misturar um conjunto de soluções de precipitantes, aditivos e soluções tampão com a solução de proteína para promover a formação da fase cristalina [4], [5]. Todavia, diversos parâmetros físicos não podem ser controlados durante esses ensaios e amostras cristalinas são obtidas por tentativa e erro e com alta taxa de irreprodutibilidade.

Uma vez obtidas amostras monocristalinas, experimentos de difração de raios X são realizados. Os dados de difração, processados e trabalhados, permitem a construção de um modelo para a estrutura tridimensional da macromolécula. De posse desse modelo estrutural, são feitas as análises da estrutura proteica, dos sítios ativos e das principais interações intermoleculares das quais a proteína participa. O conhecimento da estrutura tridimensional das proteínas permite o estudo da relação entre função e forma, e estudos da dinâmica das mudanças conformacionais, induzidas ou inerentes à macromolécula, através de simulações computacionais. Por fim, conhecer detalhes estruturais possibilita o desenho de ligantes inibidores de sítio ativo e, conseqüentemente, o design de fármacos.

O processo que permite a determinação estrutural de proteínas através da cristalografia, desde a obtenção de proteínas até as análises estruturais finais, está ilustrado na Figura 1.1.1. O conhecimento envolvido em cada uma dessas etapas é de fundamental importância na formação do profissional da área, que, ainda que se especialize, necessitará de uma compreensão geral do processo. O trabalho registrado nessa dissertação engloba o estudo de três proteínas

distintas: a esfingomielinase D (SMase D), o complexo NahK/NahL e o domínio E2 do papiloma vírus bovino tipo 1 (E2 do BPV-1). O estudo desses diferentes sistemas garantiu que todas as etapas do processo fossem visitadas, conforme esquematizado na Figura 1.1.1.

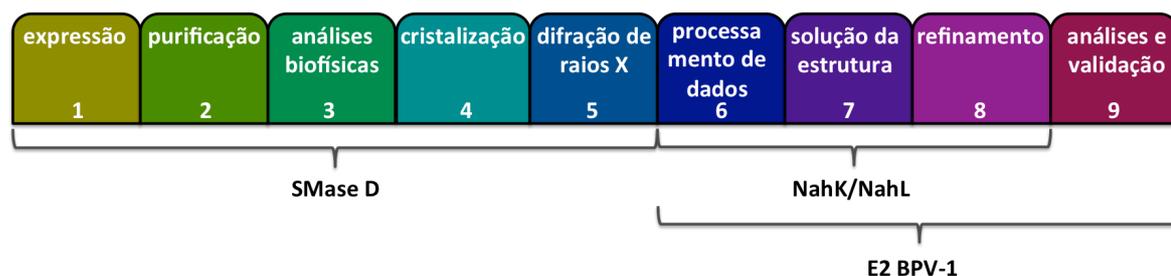


Figura 1.1.1 – Etapas sequenciais envolvidas no processo de determinação de estruturas de proteínas por cristalografia, partindo da expressão por um vetor. O trabalho com a SMase D envolveu as etapas de 1 a 5, com o complexo NahK/NahL, de 6 a 8, e com E2 de BPV-1, de 6 a 9.

Para a realização desse trabalho, contamos com a infraestrutura do laboratório BioEst do Instituto de Ciências Biológicas (ICB) da UFMG, na realização das etapas 1-3, com o Laboratório Nacional de Biociências (LNBio), na etapa 4, com o Laboratório Nacional de Luz Síncrotron, na etapa 5, e com o Laboratório de Cristalografia (LabCri) do Instituto de Ciências Exatas (ICEx) da UFMG, nas etapas 6-9.

1.2 Objetivos

Objetivo geral

Executar todas as etapas envolvidas no processo de determinação de estruturas de proteínas por cristalografia, partindo da expressão proteica até a execução das análises necessárias para a validação do modelo proposto, seguindo a série de etapas ilustrada na Figura 1.1.1.

Objetivos específicos

- Expressar, purificar e cristalizar a fosfolipase D de *Corynebacterium ulcerans* (SMase D);
- Descrever a estrutura da proteína E2 do papiloma vírus bovino tipo 1;
- Descrever a estrutura do complexo NahK/NahL de *Pseudomonas putida*.

Materiais e Métodos

2.1 Proteínas

Pequenas máquinas biológicas

A maior parte do maquinário celular é composto por proteínas e cadeias de ácidos nucleicos, sendo proteínas as responsáveis pela execução da maior parte dos processos complexos que possibilitam a vida. Elas podem ser classificadas conforme sua função como anticorpos, enzimas, componentes estruturais, proteínas mensageiras, transportadoras e até mesmo de armazenamento.

A estrutura de cada uma dessas macromoléculas codifica sua dinâmica conformacional que por sua vez está intimamente ligada com sua função. O entendimento da conexão entre a estrutura tridimensional das proteínas e sua dinâmica é, portanto, crucial para compreender o seu funcionamento. [6], [7].

Dois marcos na história da bioquímica estrutural foram a determinação estrutural do DNA por Watson e Crick em 1955 [8] e a determinação estrutural da mioglobina por Kendrew *et al.* em 1958 [9] possibilitados por resultados de experimentos de difração de raios X por cristais.

Atualmente contamos com inúmeras técnicas de bioquímica e robustos métodos computacionais que permitem estudos de dinâmica, função, estrutura e suas correlações. Os avanços nos estudos das proteínas se refletem em números nos bancos de dados disponíveis, como o *Protein Data Bank* (PDB) [3], onde novas estruturas de macromoléculas biológicas são depositadas e que hoje contém

aproximadamente 128500¹ estruturas determinadas por técnicas como difração de raios X, ressonância nuclear magnética (NMR), microscopia eletrônica e difração de nêutrons. Do PDB, os domínios estruturais proteicos são agrupados em superfamílias, quando há evidência de um ancestral comum, e hoje é possível prever a classificação de domínios estruturais a partir da sequência de aminoácidos graças a bancos como o CATH-Gene 3D [10], [11], que conta com 53 milhões de domínios proteicos classificados em 2737 superfamílias.

Composição e estrutura de proteínas

Fundamentalmente, proteínas são polímeros cujas unidades fundamentais são aminoácidos. Esses são moléculas caracterizadas por um grupamento amina (NH_2) e um grupo carboxila (COOH) ligados a um único carbono (C_α) que completa sua esfera de coordenação com um átomo de hidrogênio e um grupo radical, chamado cadeia lateral (R), como mostra a Figura 2.1.1. Em proteínas de organismos eucariotos encontramos majoritariamente aminoácidos do tipo L² (Figura 2.1.1) com vinte tipos distintos de radicais, ou cadeias laterais, que variam em tamanho, forma e propriedades químicas (Apêndice A).

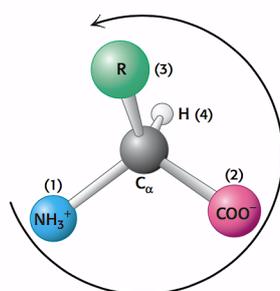


Figura 2.1.1 – Diagrama esquemático de um L aminoácido em pH neutro. A sequência no sentido anti-horário vai do substituinte com maior prioridade para o de menor prioridade. O centro quiral é denominado C_α .

¹ Com 40540 sequências proteicas distintas, sendo que organismos complexos codificam entre 20 e 30 mil proteínas diferentes.

² O carbono alpha nos aminoácidos é um centro quiral, o que confere a estes atividade óptica. Existem duas conformações possíveis: os estereoisômeros L e D, que rotacionam luz plana polarizada para a esquerda e direita respectivamente.

Cada aminoácido é codificado por um ou mais trios de bases nitrogenadas ordenadas, chamados de códon (Apêndice A). Uma sequência de aminoácidos, ou cadeia polipeptídica, é portanto especificada por uma sequência de códon no DNA, transcrita para um RNA mensageiro. Os ribossomos, compostos por proteínas e RNA ribossômico, constituem a complexa maquinaria responsável pela síntese proteica nas células. Eles leem e traduzem a informação genética contida no RNA mensageiro num processo chamado tradução e recrutam aminoácidos [2] que se conectam covalentemente numa ligação planar entre os grupamentos amina e carboxilato em uma reação do tipo ácido-base, ilustrada na Figura 2.1.2, construindo gradualmente a cadeia polipeptídica. Os aminoácidos em sequência constituem a estrutura primária da proteína. É a natureza das cadeias laterais dos aminoácidos e a interação deles com o ambiente celular que determina a forma de seu enovelamento [12]-[14].

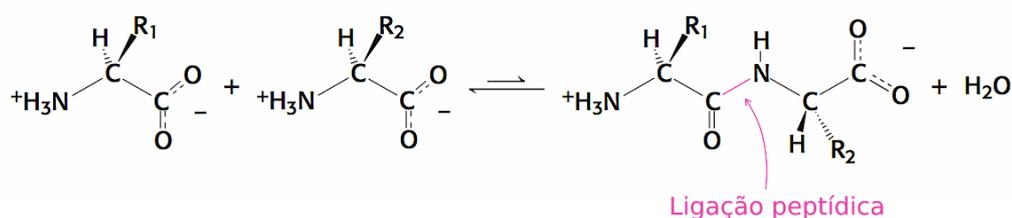


Figura 2.1.2 – Formação da ligação peptídica.

As cadeias de aminoácidos são estruturas flexíveis, entretanto, existem algumas restrições conformacionais. A ligação peptídica possui natureza planar graças à ressonância com a ligação dupla C=O, e impõe que o carbono e o nitrogênio envolvidos na ligação, o oxigênio do grupo C=O, o hidrogênio do grupo N-H e os dois C_α permaneçam num mesmo plano (Figura 2.1.3), restringindo a ligação a duas possíveis configurações: *trans* e *cis*. Contrariamente, as ligações entre o grupo amina e o C_α, e entre o C_α e o grupo carboxila são ligações simples e permitem a rotação dos grupos rígidos adjacentes. A rotação em torno delas é especificada pelos dois ângulos phi (φ) – entre o N-H e o C_α – e psi (ψ) – entre o C_α e o C=O, mas esses também são limitados em função do impedimento estérico entre átomos. Os valores permitidos para os ângulos φ e ψ foram

esquemáticos no diagrama que recebeu o nome de gráfico de Ramachandran (Apêndice A).

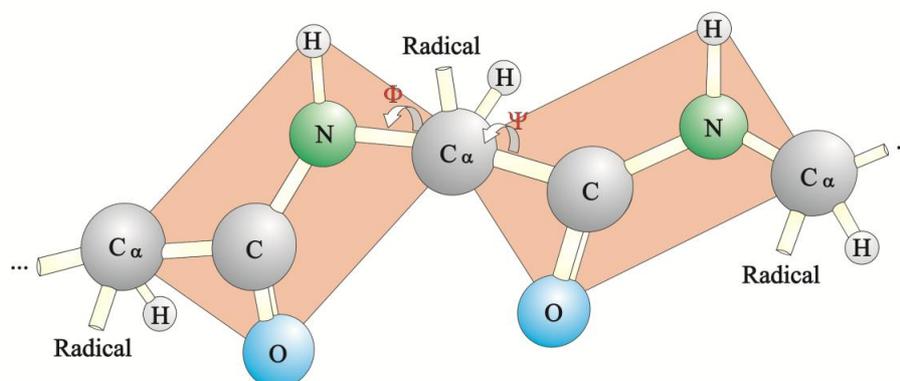


Figura 2.1.3 - Definição dos ângulos psi e phi na ligação peptídica

Com as restrições na disposição dos aminoácidos em cadeia, o enovelamento proteico ocorre localmente seguindo padrões conformacionais classificados como estrutura secundária que se mantém principalmente por ligações de hidrogênio entre aminoácidos distintos. Os elementos de estrutura secundária mais comuns são as chamadas α -hélices e folhas- β mostradas na Figura 2.1.4.

A estrutura terciária se refere, por sua vez, à organização espacial tridimensional de todos os átomos da proteína, incluindo interações entre aminoácidos distantes na estrutura primária e entre elementos de estrutura secundária. Ela se mantém principalmente devido à presença de inúmeras interações fracas como pontes de hidrogênio e interações de Van der Waals, mas também estão presentes ligações covalentes como pontes dissulfeto, comuns entre aminoácidos do tipo cisteínas.

Por fim, a interação entre as subunidades de proteínas multiméricas ou de grandes agregados proteicos constitui a estrutura quaternária. As mais variadas funções atribuídas às proteínas são executadas por estruturas que variam de monômeros isolados a estruturas com centenas de unidades [1], [15].

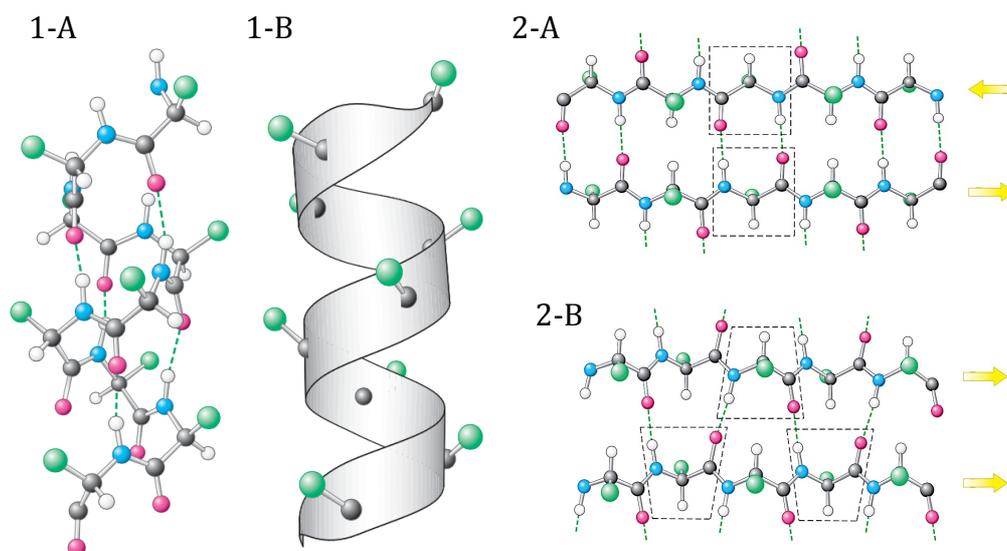


Figura 2.1.4 - Diagrama estrutural dos elementos de estrutura secundária mostrando C_{α} (preto), cadeias laterais (verdes), átomos de oxigênio (vermelho), átomos de nitrogênio (azul) e hidrogênios (branco). 1) Diagrama estrutural das α -hélices: A) representação lateral mostrando ligações de hidrogênio entre os grupos NH e CO em pontilhado verde; B) representação em ribbon da hélice com as cadeias laterais apontando para fora. 2) Diagrama estrutural das folhas- β : A) fitas adjacentes em direções opostas; B) e na mesma direção. O pontilhado verde mostra as ligações de hidrogênio entre os grupos NH e CO.

As figuras utilizadas nessa seção foram retiradas do livro: "Biochemistry", por Berg, Jeremy M., John L. Tymoczko, and Lubert Stryer.; (2002): 1515.

2.2 Produção e purificação de proteínas e análises

Expressão de proteínas recombinantes

A maior parte das proteínas investigadas atualmente são obtidas em larga escala através da expressão dos genes que as codificam por organismos mais simples, o que é possível graças à tecnologia de DNA recombinante [16]. A sequência de bases nitrogenadas que codifica a proteína de interesse é obtida através de estudos bioquímicos e de genética molecular. Uma vez dispondo do gene, ele deve ser amplificado para então ser clonado num vetor de expressão que será recebido pelo organismo responsável pela expressão, também chamado de sistema hospedeiro, *e.g.* bactérias e leveduras. O vetor de expressão é

usualmente um plasmídeo, ou seja, um pedaço de DNA circular que pode ser inserido no sistema hospedeiro. Ele é desenhado e planejado para receber o gene a ser expresso na orientação correta, para ter a expressão ativada por um indutor (comumente IPTG³) e para conferir resistência a algum antibiótico específico. Também é possível incluir uma pequena sequência de aminoácidos (*tag*) que se associará à proteína de interesse, que pode, por exemplo, facilitar a purificação da proteína em colunas de afinidade ou a sua clivagem por certas enzimas ⁴.

Os organismos hospedeiros recebem o vetor de expressão num processo chamado transformação e são posteriormente selecionados pela resistência ao antibiótico codificado no vetor. Os organismos podem então ser cultivados em meios de cultura que contém nutrientes e outras substâncias que propiciam sua replicação. Quando dispendo de quantidade apropriada de organismos, a indução da expressão é feita e a cultura é mantida na temperatura mais adequada e em constante agitação. Para confirmar a expressão, é feita a lise das células e a separação das frações solúveis e insolúveis para análise via eletroforese em gel de poliacrilamida e dodecil-sulfato de sódio (SDS). As posições de bandas no gel de poliacrilamida-SDS, correspondentes às proteínas presentes na fração solúvel, são comparadas com as posições de marcadores de peso molecular conhecidos. A presença de uma banda cujo peso molecular corresponde ao peso da proteína em estudo indica êxito na expressão. O próximo passo é a purificação do conteúdo celular para confirmação da presença da proteína em questão e obtenção de proteína pura para outras análises e ensaios [18].

É importante ressaltar que os protocolos com os parâmetros adequados que favorecem a replicação celular e a produção de proteína são desenvolvidos para os sistemas individualmente e requerem, muitas vezes, um longo trabalho que envolve testes e variações de parâmetros diversos.

³ Isopropil- β -D-tiogalactosídeo

⁴ Os processos de isolamento e caracterização de genes e de clonagem vão além do escopo desse trabalho. Para mais detalhes, ver bibliografia [15],[1] e [17].

Cromatografia por afinidade por cauda de histidinas

A cromatografia por afinidade separa moléculas pela afinidade que elas têm por uma matriz cuja composição é configurada para o experimento e pode ser qualquer contanto que insolúvel e porosa. Uma coluna da matriz é caracterizada pela presença do material pelo qual a macromolécula de interesse possui grande afinidade ligado à resina (ligante preso). Quando a solução que contém a proteína de interesse passa pela coluna, ela se liga especificamente ao ligante e os contaminantes são lavados. Para desprender a proteína da matriz utiliza-se um ligante livre que com ela compete por ter maior afinidade pelo ligante preso.

Como mencionado anteriormente, a produção de proteínas por técnicas que utilizam DNA recombinante em bactéria, levedura ou outros organismos permitiu a inclusão de elementos que auxiliam na purificação direto na estrutura primária por design genético. Dentre as modificações mais comumente empregadas, destaca-se a adição de uma sequência de Histidinas (*His-tag*) presa ao N ou C-terminal da proteína recombinante. Isso mostrou afetar pouco o enovelamento proteico e permite isolar a macromolécula de interesse utilizando uma coluna cromatográfica que contenha cobre ou níquel como ligantes presos, sendo a última uma coluna comercial [19]. Para remover a proteína com *His-tag* da coluna utiliza-se um tampão contendo imidazol que é misturado ao tampão inicial tal que haja um gradiente de concentração de imidazol na coluna e a eluição da macromolécula aconteça em um momento característico. A absorção de luz UV pelo eluente da coluna é monitorada em tempo real uma vez que os aminoácidos tirosina e triptofano absorvem luz UV de 280nm de forma específica [20] e permitem a identificação da presença de cadeias de aminoácidos nas frações de eluentes que são coletadas. O gradiente de concentração de imidazol deve ser suave e o volume das frações deve ser pequeno o suficiente para permitir a separação de porções distintas correspondentes a grupos de moléculas com afinidades distintas pela matriz [21]. O processo de cromatografia por afinidade por um ligante está ilustrado na Figura 2.2.1.

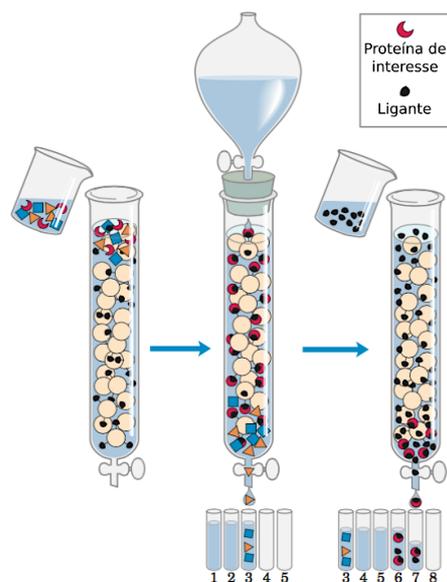


Figura 2.2.1 – Cromatografia por afinidade a um ligante. A solução com compostos diversos incluindo a proteína de interesse (meia lua vermelha) passa pela coluna composta pela resina (esferas maiores) e o ligante preso (esferas menores). A proteína se prende ao ligante, enquanto os demais compostos iniciais eluem pela coluna. Por fim, o ligante livre que compete com a proteína pela afinidade à coluna é adicionado e a proteína elui. O eluato é coletado em frações. (Imagem adaptada de [15])

Cromatografia por exclusão de tamanho

A cromatografia por exclusão de tamanho separa moléculas presentes na solução com base no seu peso molecular e na sua forma, utilizando uma matriz composta por esferas microscópicas de agarose e dextrano reticulados [20] , [21]. No nível microscópico essas esferas são perfuradas por uma rede de canais cujo diâmetro médio é controlado na produção do material de forma a determinar o tamanho das partículas que podem penetrá-los. A resolução da matriz é determinada pela largura da distribuição dos diâmetros de seus canais em torno do valor médio.

As proteínas passam por uma coluna dessa matriz impelidas pela gravidade ou por pressão aplicada. Aquelas de menor tamanho penetrarão também os canais mais estreitos e a velocidade com que elas passam de um lado a outro da coluna será reduzido devido ao tempo que elas levam migrando no interior das esferas. Enquanto isso as moléculas maiores vão passar por fora de várias das esferas, pois os diâmetros dos canais restringem seu percurso, fazendo com que elas passem mais rapidamente pela matriz.

Novamente, o eluente da coluna é monitorado com medidas de absorção de luz UV e frações dele são coletadas em volumes pequenos o suficiente para separação das partes que correspondem a proteína de interesse ou não. O método é eficiente na separação de macromoléculas cujo peso molecular difira em até 15% na média e tem como vantagens a possibilidade de manter o meio em que a proteína se encontra estável e a reprodutibilidade [21]. O processo de cromatografia por exclusão de tamanho está ilustrado na Figura 2.2.2.

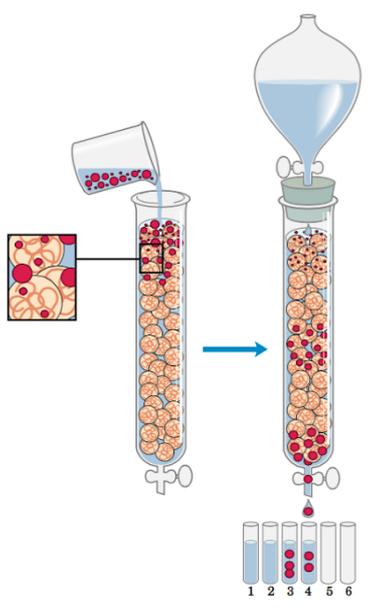


Figura 2.2.2 – Cromatografia por exclusão de tamanho. A solução com compostos de diversos tamanhos, incluindo a proteína de interesse, passa pela coluna composta pela resina (esferas porosas). O tempo de eluição pela coluna é inversamente proporcional ao tamanho da partícula. O eluato é coletado em frações possibilitando a separação por tamanho. (Imagem adaptada de [15])

Espalhamento dinâmico de luz – DLS

O espalhamento dinâmico de luz (Dynamic Light Scattering – DLS) é uma técnica espectroscópica utilizada na análise de flutuações na intensidade da luz espalhada por partículas em suspensão que permite a determinação da evolução temporal da distribuição de raios hidrodinâmicos médios dos espalhadores e outras propriedades dinâmicas.

No experimento de DLS, a luz de um laser passa por um polarizador e em seguida pela amostra, onde o campo elétrico da luz incidente induz uma polarização

oscilante nos elétrons das moléculas, que portanto se tornam fontes secundárias de radiação. Nesse processo a radiação será reemitida com mesma frequência da luz incidente e, portanto, ele pode ser interpretado simplesmente como um espalhamento elástico. A luz espalhada pela amostra passa por um segundo polarizador e entra em um detector cuja posição angular (θ) em relação à direção do feixe incidente determina um vetor \mathbf{q} na direção do espalhamento. O movimento aleatório (Browniano) devido às interações térmicas entre as moléculas que compõem a amostra faz com que o campo elétrico espalhado flutue no tempo também de forma aleatória. A intensidade espalhada é medida para uma gama de ângulos θ durante um período de tempo T suficientemente grande. Considerando que todos os estados do sistema analisado sejam igualmente prováveis, ou seja, que o sistema é ergódico, a intensidade da luz espalhada dependente do tempo $I(q, t)$ flutua em torno de um valor médio dado por:

$$\langle I(q, t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T I(q, t) dt \quad 2.2.1$$

onde a magnitude q do vetor de espalhamento \mathbf{q} é dada por:

$$q = \frac{4\pi n}{\lambda} \sin\left(\frac{\theta}{2}\right) \quad 2.2.2$$

onde n é o índice de refração da amostra e λ é o comprimento de onda da luz incidente.

Para dois instantes de tempo t e $t + \tau$, a intensidade da luz espalhada para um determinado \mathbf{q} varia pouco se τ tende a zero, mas pode apresentar uma grande variação para um valor de τ grande. A relação entre as intensidades medidas nos dois instantes é fornecida pela função de autocorrelação:

$$C(q, \tau) = \langle I(q, t)I(q, t + \tau) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T I(q, t)I(q, t + \tau) dt \quad 2.2.3$$

Como a intensidade da luz espalhada é uma propriedade não conservativa e não periódica, sua função de autocorrelação possui os limites:

$$\lim_{\tau \rightarrow 0} \langle I(q, t) I(q, t + \tau) \rangle = \langle I(q, t)^2 \rangle \quad 2.2. 4$$

$$\lim_{\tau \rightarrow \infty} \langle I(q, t) I(q, t + \tau) \rangle = \langle I(q, t) \rangle^2$$

ou seja, a correlação entre os valores medidos para a intensidade espalhada decai monotonicamente à medida que o intervalo de tempo entre as medidas τ vai para infinito (Figura 2.2.3) e portanto vale a relação de Siegert:

$$\frac{C(q, \tau)}{\langle I(q, t) \rangle^2} = 1 + \xi |g(q, \tau)|^2 \quad 2.2. 5$$

onde ξ é uma constante do instrumento e $g(q, \tau)$ é a função de correlação normalizada do campo elétrico, exceto quando o número de espalhadores é muito pequeno ou quando o movimento deles é limitado.

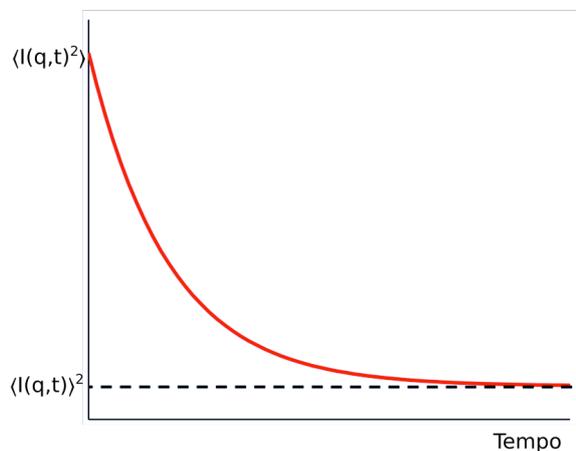


Figura 2.2.3 – Gráfico ilustrativo da função de autocorrelação temporal $C(q, \tau)$

Para espalhadores esféricos monodispersos, $g(q, \tau)$ é dada por:

$$g(q, \tau) = e^{(-D_f q^2 \tau)} \quad 2.2. 6$$

sendo D_f o coeficiente de difusão das partículas dado pela equação de Stokes-Einstein para partículas esféricas:

$$D_f = \frac{k_B T}{6\pi\eta R_H} \quad 2.2. 7$$

onde k_B é a constante de Boltzmann, T é temperatura da amostra, η é a viscosidade do meio e R_H é o raio hidrodinâmico das partículas.

Conhecendo-se o comprimento de onda da luz incidente, o ângulo do espalhamento e o índice de refração do meio, é possível determinar a magnitude do vetor \mathbf{q} pela equação 2.2. 2. As medidas da intensidade da luz espalhada permitem o cálculo de $g(q, \tau)$ através da relação de Siegert para sistemas monodispersos e cujos espalhadores são esféricos. Finalmente, é possível calcular o coeficiente de difusão D_f através do ajuste das curvas 2.2. 6, o que nos fornece o raio hidrodinâmico das partículas pela equação de Stokes-Einstein (equação 2.2. 7).

Na análise da proteína em solução, após as etapas de purificação, a medida de DLS nos informa a distribuição dos tamanhos das partículas presentes e, portanto, seu estado de dispersão, que deve ser preferencialmente monodispersa para ensaios de cristalização [24]. A partir dos tamanhos aproximados das partículas na solução, também é possível estimar a quantidade de monômeros que compõem a unidade proteica estável em solução [24],[26],[27].

Dicroísmo circular

Luz linearmente polarizada é composta por duas componentes de mesma amplitude e frequência, circularmente polarizadas para a direita (R) e para a esquerda (L), respectivamente. A espectroscopia de dicroísmo circular (CD) avalia a diferença na absorção de luz circularmente polarizada pelo material de interesse. Se ambas componentes (R e L) não são absorvidas, ou são absorvidas da mesma forma, a polarização inicial é reconstituída após a passagem pela

amostra. Entretanto, se as componentes R e L forem absorvidas de forma distinta, a luz que sai da amostra apresentará uma polarização elíptica. A diferença ΔA entre a absorbância das componentes R e L é dada em termos da elipticidade por:

$$\theta = 32.98 \Delta A \quad 2.3.1$$

sendo $\theta = \tan^{-1}(b/a)$ onde b é o raio menor da elipse e a é o raio maior [28].

O sinal de dicroísmo circular será observado sempre que o cromóforo for quiral, seja a quiralidade intrínseca à sua estrutura, por estar ligado covalentemente a um centro quiral na molécula, ou por estar em um local assimétrico devido à estrutura tridimensional da molécula. Em proteínas, os cromóforos de interesse incluem: a ligação peptídica (absorve abaixo de 240nm), cadeias laterais de aminoácidos aromáticos (absorvem entre 260 e 320nm) e pontes dissulfeto (bandas de absorção largas centradas próximo de 260nm) [24]–[26]. Dessa forma, é possível perceber a assinatura de estruturas secundárias e terciárias no sinal de CD, e portanto podemos estimar a composição estrutural com a ajuda de algoritmos tais como o CDNN [31], SELCON [32] e CONTIN [29], [34].

Além de informações estruturais, o espectro de CD avaliado em função da temperatura permite a análise do processo de desenovelamento e enovelamento proteico. Quando as proteínas desenovelam, elas perdem suas estruturas secundárias e as bandas de CD mudam. Dessa forma é possível avaliar a termodinâmica do processo monitorando a elipticidade para um comprimento de onda ou coletando o espectro completo em função da temperatura [26], [31]–[33].

2.3 Cristalização

Nucleação e crescimento de cristais

Podemos descrever um cristal como um sólido homogêneo e anisotrópico constituído microscopicamente por uma repetição ordenada nas três direções do espaço de seus constituintes. É o arranjo mais simples no qual os átomos se organizam formando um sólido macroscópico e que, para uma infinidade de compostos, é a configuração do estado de mais baixa energia [37].

Para que a cristalização aconteça a partir de uma solução, é necessário levar o sistema a um estado de não equilíbrio no qual a energia livre de Gibbs do sólido G_s a ser formado e da solução G_L sejam distintas [38]. Se a diferença $\Delta G = G_L - G_s < 0$, é provável que aconteça a formação de um sólido. Particularmente, a cristalização de proteínas requer a criação de um estado de não equilíbrio no qual a concentração da macromolécula excede o limite de solubilidade (supersaturação) sobre certas condições termodinâmicas. O equilíbrio é reestabelecido com a formação de material sólido, eventualmente cristalino, à medida que o limite de saturação da solução contendo as proteínas é atingido. Esse processo está ilustrado na Figura 2.3.1.

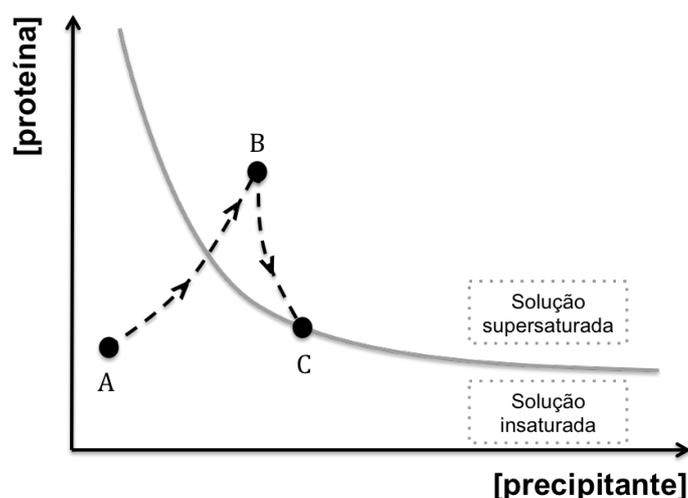


Figura 2.3.1 – Diagrama de fase representando a cristalização de proteínas. No processo $A \rightarrow B$, a supersaturação é atingida. De $B \rightarrow C$, acontece a formação do sólido, reduzindo a concentração de proteína em solução até que o limite de solubilidade é atingido (ponto C)

A cristalização de um certo material envolve o controle de inúmeros parâmetros e a formação inicial do sólido é apenas um possível primeiro passo, uma vez que pode haver a formação tanto de um precipitado amorfo quanto de um sólido que pode ou não ser cristalino. Para facilitar a análise, dividimos o processo de cristalização em três fases distintas: nucleação, crescimento e fim do crescimento [39]. A nucleação, ou formação de um núcleo crítico, é a parte mais complexa da cristalização e consiste numa transição descontínua de fase no qual moléculas passam de um estado completamente desordenado para um estado ordenado [40]. A formação de um núcleo crítico é um processo de probabilidade baixa e, de acordo com a teoria clássica de nucleação [41], a taxa de formação de núcleos críticos τ é proporcional a densidade de moléculas (ρ) e à exponencial negativa do custo energético (ΔG) de sua formação:

$$\tau \propto \rho e^{(-\Delta G/k_B T)}. \quad 2.3.1$$

onde k_B é a constante de Boltzmann e T é a temperatura do sistema. Uma vez formado o núcleo, o crescimento do cristal envolve o transporte do material em solução até a superfície do cristal e a associação das moléculas à rede cristalina. A probabilidade de uma dada molécula se ligar ao cristal (p) é determinada pela razão entre a taxa de crescimento do cristal (k) e a taxa de colisões de moléculas com o cristal por unidade de concentração e por unidade de área (Z).

A taxa de crescimento do cristal pode ser obtida experimentalmente⁵ [39], e a taxa de colisões $Z = \sqrt{(2\pi m/k_B T)}$, onde m é a massa do cristal, é obtida pela teoria cinética [42]. Para a lisozima, a probabilidade da molécula se ligar ao cristal ($p = k/Z$), é da ordem de 10^{-7} .

O fim do crescimento do cristal acontece quando a solução se torna instaurada, *i.e.* quando a concentração de proteína em solução se torna inferior ao seu limite de solubilidade (solução insaturada), e o equilíbrio é reestabelecido. Na prática, contudo, o crescimento dos cristais cessa ainda que o equilíbrio não tenha se estabelecido. Esta última etapa da cristalização ainda não é bem compreendida e

⁵ Para a um experimento com cristais de lisozima, $k = 5 \times 10^{-5}$ cm/s

uma das hipóteses é que outros mecanismos levam ao fim do processo de crescimento, como, por exemplo, o acúmulo de erros na agregação das moléculas de forma a tornar novas ligações desfavoráveis energeticamente.

Métodos de cristalização de proteínas

A maior parte dos métodos utilizados na cristalização de proteínas consistem fundamentalmente na redução da solubilidade da proteína levando à supersaturação da solução. Isso pode ser feito de inúmeras formas, tais como: mudança de pH, adição de íons ou polímeros e mudança da temperatura [19],[40], [41].

Uma técnica amplamente utilizada na cristalização de proteínas é a difusão de vapor, no qual a perda de vapor leva à supersaturação (processo $A \rightarrow B$ na Figura 2.3.1). O aparato experimental consiste em um poço da solução contendo agentes precipitantes e uma gota de 1-10 μ l que contém solução de proteína e solução de poço, comumente numa razão de 1:1. Os dois principais métodos de cristalização de proteínas por difusão de vapor, *Hanging drop* (gota pendurada) e *Sitting drop* (gota sentada) estão ilustrados na Figura 2.3.2.

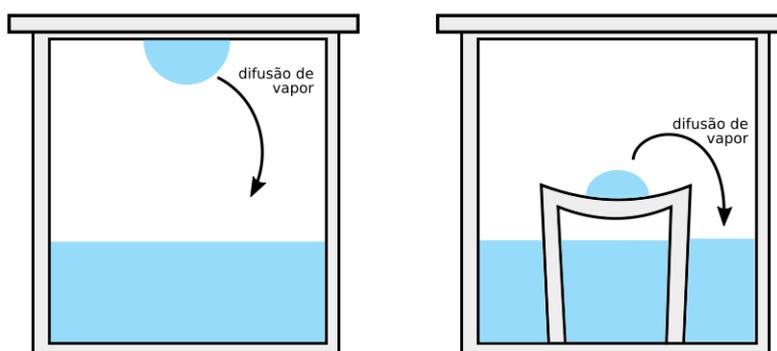


Figura 2.3.2– Diagrama do aparato experimental dos métodos de *Hanging drop* (esquerda) e *Sitting drop* (direita).

Triagem e otimização de condições de cristalização

O processo de obtenção de cristais de proteína adequados para experimentos de difração de raios X se inicia com a busca pelas condições químicas, bioquímicas e físicas que propiciam a formação de algum material cristalino proteico. Posteriormente, alterações sistemáticas e incrementais dessas condições são feitas na tentativa de otimizar os parâmetros até que cristais de tamanho adequado para os experimentos sejam obtidos. Duas abordagens são utilizadas corriqueiramente: (i) o método de matriz esparsa [45] baseado na variação sistemática de parâmetros importantes como pH, temperatura, o tipo do agente precipitante e sua concentração de forma a varrer um grande número de condições; (ii) e o método de *ShotGun* [46], que reduz o número de condições testadas de forma racionalizada, evitando a redundância entre condições presentes na varredura mais extensa. Independente da abordagem utilizada, o processo de busca da condição ótima para cristalização é, muitas vezes, uma etapa demorada uma vez que os parâmetros envolvidos podem estar correlacionados. Os avanços tecnológicos em áreas como a robótica e fluidos possibilitaram a automação da busca preliminar por condições de cristalização, de forma que alguns miligramas de proteína hoje são suficientes para testar centenas de condições de cristalização.

2.4 Teoria de espalhamento de raios X

Interação dos raios X com a matéria

O uso de radiação eletromagnética é uma das mais importantes formas de se investigar a estrutura e a dinâmica da matéria. Fenômenos de reflexão, refração, absorção, fluorescência, espalhamento e difração permitem investigar inúmeras propriedades de diversos tipos de materiais. Técnicas de caracterização comumente empregadas permitem desde a obtenção de um espectro de energia até a determinação da estrutura tridimensional de moléculas e compostos,

incluindo entidades complexas como macromoléculas biológicas. Cada região do espectro eletromagnético interage com a matéria de forma particular sendo, portanto, o comprimento de onda da radiação utilizada em cada técnica o fator determinante dos resultados. Em termos de resolução óptica, o critério de Rayleigh especifica que a separação angular mínima entre duas fontes de luz com comprimento de onda λ para que elas sejam resolvidas como dois objetos distintos deve ser:

$$\theta_{min} = \frac{\lambda}{a} \quad 2.4.1$$

onde a é a espessura da fenda e $\lambda \sim a$ [47]. O comprimento de onda da radiação X é compatível com as distâncias interatômicas típicas sendo, portanto, a radiação utilizada em experimentos que buscam resolução atômica.

Fótons de raios X ao interagir com a matéria podem ser absorvidos ou espalhados [47]. A absorção de fótons pela amostra excita elétrons para níveis mais altos de energia e pode levar à ocorrência dos fenômenos de emissão fotoelétrica, fluorescência e emissão Auger, que apenas puderam ser descritos com a emergência da mecânica quântica. Quando um fóton de raios X é absorvido por um átomo, pode haver energia suficiente para ejetar um elétron deixando o átomo ionizado num fenômeno chamado emissão fotoelétrica. Forma-se uma vacância na camada eletrônica à qual esse elétron pertencia e, se essa for uma camada mais interna do átomo, essa vacância será preenchida por um elétron pertencente a uma camada mais externa liberando energia igual à diferença energética entre as camadas de origem (externa) e destino (interna). Essa energia pode ser liberada em forma de um fóton no fenômeno de fluorescência ou ainda pode ser usada para ejetar outro elétron de uma camada mais externa, num fenômeno chamado emissão Auger (Figura 2.4.1).

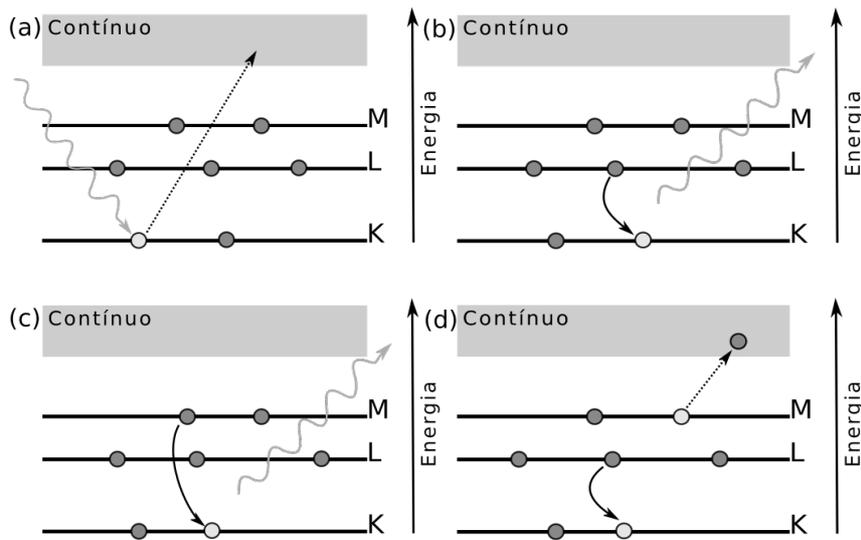


Figura 2.4.1 – Fenômenos que envolvem a absorção de raios X; (a) absorção fotoelétrica; (b) e (c) fluorescência devido à transição de um elétron da camada L e da camada M para a camada K, cujas linhas de emissão levam os nomes K_{α} e K_{β} respectivamente; (d) emissão Auger

Os fenômenos de espalhamento acontecem de forma elástica e inelástica, o que será detalhado nas próximas seções. Na descrição clássica dos fenômenos de espalhamento, o campo elétrico de uma onda eletromagnética incidente num alvo espalhador exerce uma força nos elétrons dos átomos acelerando-os de forma que eles irradie na mesma frequência da onda incidente e o espalhamento é necessariamente elástico. Para explicar a existência de ambos espalhamentos, elástico e inelástico, é necessária a descrição quântica na qual fótons incidem no alvo espalhador com momento $\hbar\mathbf{k}$ e energia $\hbar\omega$. Na interação dos fótons com o alvo, energia e momento do fóton podem ser transferidos ao espalhador resultando num fóton espalhado cuja frequência é igual (elástico) ou menor (inelástico) à do fóton incidente.

Espalhamento inelástico

O espalhamento inelástico, ou Compton, pode ser descrito em termos da colisão entre um fóton (incidente) e um elétron (alvo). Nesse processo, o fóton cede parte de sua energia para o elétron e é desviado de sua trajetória. O processo está ilustrado na Figura 2.4.2, onde θ é o ângulo entre a direção do fóton incidente e a direção do fóton espalhado. O espalhamento Compton é um

processo incoerente, ou seja, não envolve nenhuma relação de fase entre a radiação incidente e espalhada, e a perda de energia do fóton pode ser calculada considerando a conservação do momento e da energia na colisão. A variação no comprimento de onda do fóton espalhado no espalhamento Compton é dada por:

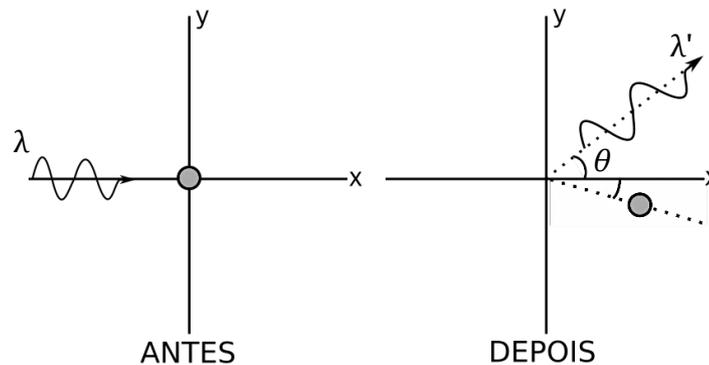


Figura 2.4.2 - Diagrama da colisão entre fóton e elétron representando o Espalhamento Compton.

$$\Delta\lambda = (\lambda' - \lambda) = \frac{h}{m_e c} (1 - \cos\theta) \quad 2.4. 2$$

onde h é a constante de Planck, m_e é a massa do elétron e c é a velocidade da luz.

Espalhamento elástico

Consideramos uma partícula de carga e e massa m na origem do sistema de coordenadas cartesianas e uma onda plana eletromagnética monocromática propagando-se ao longo do eixo x cujo campo elétrico é dado por:

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k}_i \cdot \mathbf{r} - \omega t)} \quad 2.4. 3$$

onde ω é a frequência angular, \mathbf{k}_i é o vetor de onda incidente de módulo $2\pi/\lambda$ e \mathbf{r} é o vetor que define o ponto Q de observação. Segundo a teoria de espalhamento de Thomson, esse campo exerce uma força $\mathbf{F} = e\mathbf{E}$ sobre o elétron causando a aceleração $\mathbf{a} = e\mathbf{E}/m$, tal que o elétron passa a oscilar com a mesma frequência ν

da onda incidente. A radiação emitida pela partícula carregada em movimento acelerado é da forma:

$$\mathbf{E}_d = \mathbf{E}_{0d} e^{i(\mathbf{k}_d \cdot \mathbf{r} - \omega t)} \quad 2.4.4$$

onde \mathbf{k}_d é o vetor de onda emitida de módulo $2\pi/\lambda$. Thomson mostrou que:

$$E_{0d} = \frac{1}{r} E_0 \left(\frac{e^2}{m_e c^2} \right) \sin\varphi \quad 2.4.5$$

onde m_e é a massa do elétron, e é a carga do elétron, c é a velocidade da luz, r é o módulo do vetor \mathbf{r} e φ é o ângulo entre a direção de observação e a direção de aceleração do elétron. A intensidade I da radiação emitida pelos elétrons será:

$$I = \|\mathbf{E}_{0d}\|^2 = I_i \frac{e^4}{m^2 r^2 c^4} \sin^2\varphi . \quad 2.4.6$$

que garante que podemos ignorar a contribuição dos prótons para o espalhamento uma vez que o fator $\left(\frac{e^4}{m^2}\right)$ é perto de 1837^2 vezes menor que o mesmo fator para o elétron. O espalhamento é, portanto, predominantemente feito pela nuvem de elétrons ao redor do núcleo de cada átomo. A diferença entre as intensidades total e espalhada elasticamente dá a intensidade espalhada de forma inelástica.

Espalhamento elástico por um grupo de espalhadores e interferência

Consideremos agora um grupo de N espalhadores idênticos alinhados no espaço de forma fixa num referencial fixo R :

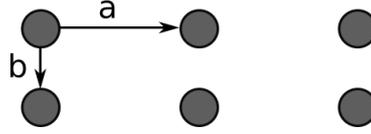


Figura 2.4.3 - Espalhadores alinhados numa rede bidimensional determinada pelos vetores \mathbf{a} e \mathbf{b} .

e uma onda plana incidente:

$$A(\mathbf{r}, t) = A_0 e^{i(\mathbf{k}_i \cdot \mathbf{r} - \omega t)}. \quad 2.4.7$$

A radiação total espalhada A_d na direção d é a superposição das ondas espalhadas por cada espalhador j nessa direção e é dada por:

$$A_d(\mathbf{r}, t) = e^{-i\omega t} \sum_{j=1}^N A_j e^{i(\mathbf{k}_d - \mathbf{k}_i) \cdot \mathbf{r}} \quad 2.4.8$$

onde A_j é a amplitude da onda espalhada por cada espalhador j , \mathbf{k}_d é o vetor de onda da onda espalhada e \mathbf{r} é o vetor entre o espalhador e o ponto de observação. A interferência entre as ondas espalhadas gera padrões de difração que são constantes no tempo no referencial fixo R .

A posição de máximos de interferência construtiva nos padrões de difração pode ser calculada a partir de uma análise geométrica do problema. Consideramos dois centros espalhadores consecutivos, $j = 1$ e $j = 2$ separados pelo vetor \mathbf{a} .

Vê-se que a diferença de caminho entre os feixes espalhados é:

$$\Delta l = |\mathbf{a}| [\cos(\rho) - \cos(\varphi)] \quad 2.4.9$$

e portanto a diferença de fase entre eles é $2\pi\Delta l/\lambda$. Os máximos de interferência construtiva acontecerão quando somarmos ondas em fase, ou seja: $\Delta l = n\lambda$, onde n é um número inteiro. Reescrevemos essa condição em termos dos vetores unitários \mathbf{s}_0 e \mathbf{s} , (tal que $\mathbf{s}_0 = 2\pi\mathbf{k}_i$ e $\mathbf{s} = 2\pi\mathbf{k}_d$) que determinam as direções da onda incidente e da onda espalhada respectivamente (Figura 2.4.4):

$$\mathbf{a} \cdot (\mathbf{s} - \mathbf{s}_0) = n\lambda, \quad 2.4. 10$$

que é conhecida como condição de Laue.

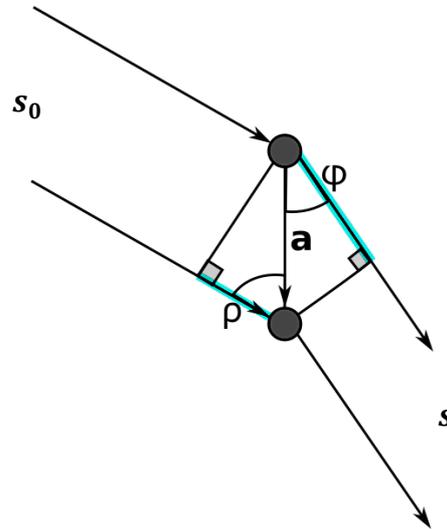


Figura 2.4.4 - Diferença de caminho óptico entre ondas espalhadas por dois espalhadores consecutivos.

Difração por cristais 3d

Formalmente, um material pode ser chamado cristal se essencialmente apresenta picos de interferência construtiva discretos em seu padrão de difração [48]. A rede formada por estes picos pode ser descrita pelos vetores:

$$\mathbf{h} = \sum_{i=1}^n h_i \mathbf{a}_i^* \quad (n \geq 3) \quad 2.4. 11$$

onde \mathbf{a}_i^* são os vetores que constituem a base do espaço recíproco e h_i são coeficientes inteiros. Se extrapolarmos a equação 2.4. 10 para uma rede tridimensional de espalhadores, teremos a condição para haver interferência construtiva, na direção \mathbf{s} da radiação espalhada por um cristal, escrita na forma:

$$\mathbf{a}_i \cdot (\mathbf{s} - \mathbf{s}_0) = h_i \lambda \quad 2.4. 12$$

onde \mathbf{a}_i para $i = 1, 2$ ou 3 são os vetores da rede periódica tridimensional, e h_i são números inteiros. Pode-se mostrar que⁶:

$$\frac{(\mathbf{s} - \mathbf{s}_0)}{\lambda} = \mathbf{h} \quad 2.4. 13$$

onde \mathbf{h} são vetores da rede recíproca e a condição de Laue para a difração (eq 2.4. 10) pode ser reescrita como:

$$\mathbf{a}_i \cdot \mathbf{h} = h_i \quad 2.4. 14$$

Uma outra interpretação do fenômeno de difração é possível se considerarmos que no interior do cristal exista uma série de planos paralelos de átomos, separados por uma distância d . Supondo o espalhamento elástico, quando incidirmos radiação de comprimento de onda λ , os ângulos de incidência θ e de reflexão serão iguais, como mostra a Figura 2.4.5.

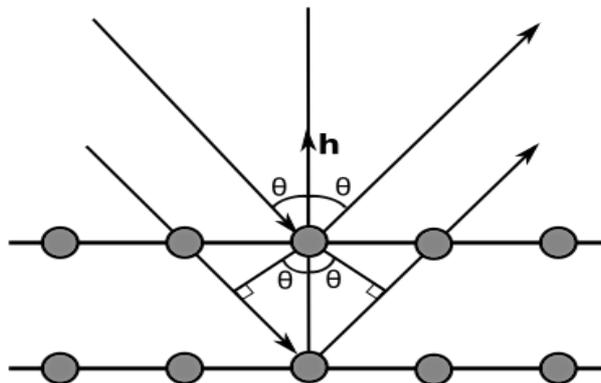


Figura 2.4.5 – Difração do ponto de vista da interpretação de Bragg.

⁶ Mais detalhes por B.E. Warren, "X-ray diffraction" [49]

Interferência construtiva será observada sempre que a diferença de caminho entre os feixes difratados seja um múltiplo inteiro de comprimento de onda. Ou seja, sempre que:

$$2d\sin\theta = n\lambda \quad 2.4. 15$$

A equação 2.4. 15 é conhecida como a lei de Bragg, que é equivalente à condição de Laue⁷. Entretanto é importante notar que a descrição geométrica acima não descreve a física do fenômeno de difração.

Para calcular a amplitude da onda difratada em um experimento de difração de raios X por cristais, consideramos apenas o espalhamento devido à nuvem de elétrons ao redor do núcleo de cada átomo. A fim de descrever a posição dos espalhadores reais é necessário definir uma função de distribuição $\rho(\mathbf{r}) = |\psi(\mathbf{r})|^2$, também chamada de função de densidade eletrônica, que dá a probabilidade de um elétron estar contido em uma unidade de volume na posição determinada pelo vetor posição \mathbf{r} . A amplitude A_T da onda espalhada por um elemento de volume $d\mathbf{r}$ do cristal será, portanto:

$$A_T \propto \rho_{crist}(\mathbf{r})d\mathbf{r}. \quad 2.4. 16$$

Para estimar a amplitude da onda espalhada por todo o cristal, consideramos inicialmente os feixes espalhados por elementos de volume separados por um vetor \mathbf{r} , conforme indicado na Figura 2.4.6.

Observamos, novamente, que a fase da onda incidente sofrerá uma variação de $\mathbf{r} \cdot (\mathbf{k}_d - \mathbf{k}_i)$, e portanto a amplitude espalhada elasticamente será dada por:

$$A_T \propto A_0 \rho_{crist}(\mathbf{r}) e^{i[(\mathbf{k}_d - \mathbf{k}_i) \cdot \mathbf{r} - \omega t]} d\mathbf{r}. \quad 2.4. 17$$

⁷ Para demonstração, ver B.E. Warren, "X-ray diffraction"

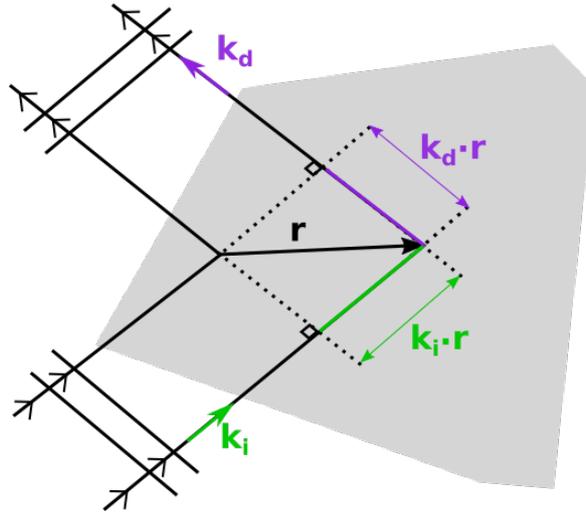


Figura 2.4.6 – Diferença de caminho óptico na difração em um cristal

A amplitude espalhada por todo o cristal contém a soma da contribuição de cada par de elemento de volume:

$$A_T \propto \int A_0 \rho_{crist}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r} - \mathbf{v}t)} d\mathbf{r} = A_0 e^{-i\omega t} \int \rho_{crist}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} \quad 2.4. 18$$

onde $\mathbf{h} = \frac{(\mathbf{k}_d - \mathbf{k}_i)}{2\pi}$ é um vetor do espaço recíproco. Vemos que a amplitude total difratada é proporcional à transformada de Fourier da densidade eletrônica:

$$A_T \propto \int \rho_{crist}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} = \mathcal{F}_r[\rho_{crist}(\mathbf{r})](\mathbf{h}) \quad 2.4. 19$$

Mas como cristais são objetos periódicos, podemos representar sua densidade eletrônica com auxílio da função $L(\mathbf{r})$, que descreve matematicamente a rede infinita tridimensional por:

$$L(\mathbf{r}) = \sum_{h,k,l=-\infty}^{+\infty} \delta(\mathbf{r} - \mathbf{r}_{n_1, n_2, n_3}) \quad 2.4. 20$$

onde $\mathbf{r} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$ e n_1, n_2 e n_3 são inteiros, $\mathbf{a}_1, \mathbf{a}_2$ e \mathbf{a}_3 são os vetores base do espaço real e δ denota a distribuição delta de Dirac. A densidade

eletrônica dentro do cristal será dada pela convolução $\rho_{crist}(\mathbf{r}) = \rho_{cel}(\mathbf{r}) * L(\mathbf{r})$.

Dessa forma, a transformada de Fourier da densidade eletrônica será:

$$\mathcal{F}[\rho_{crist}(\mathbf{r})](\mathbf{h}) = \int_{vol} \left[\rho_{cel}(\mathbf{r}) * \sum_{n_1, n_2, n_3 = -\infty}^{+\infty} \delta(\mathbf{r} - \mathbf{r}_{n_1, n_2, n_3}) \right] e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} \quad 2.4. 21$$

e pelo teorema da convolução, que diz que a transformada de Fourier da convolução de duas funções é o produto das transformadas de cada uma delas, temos que:

$$\mathcal{F}[\rho_{crist}(\mathbf{r})](\mathbf{h}) = \mathcal{F}[L(\mathbf{r})](\mathbf{h}) \cdot \mathcal{F}[\rho_{cel}(\mathbf{r})](\mathbf{h}). \quad 2.4. 22$$

Resolvendo a transformada de Fourier da distribuição delta de Dirac:

$$\mathcal{F}[\rho_{crist}(\mathbf{r})](\mathbf{h}) = \left[\sum_{n_1, n_2, n_3 = -\infty}^{+\infty} e^{2\pi i(\mathbf{h} \cdot \mathbf{r}_{n_1, n_2, n_3})} \right] \int_{cel} \rho_{cel}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} \quad 2.4. 23$$

Se considerarmos um cristal real, teremos uma rede finita e, portanto, o cristal possuirá N_1 células unitárias na direção de \mathbf{a}_1 , N_2 na direção \mathbf{a}_2 e N_3 na direção \mathbf{a}_3 . Então abrimos a exponencial do somatório entre colchetes para o cristal finito, chamada genericamente de \mathcal{S} :

$$\mathcal{S} = \sum_{n_1=1}^{N_1} e^{2\pi i n_1(\mathbf{h} \cdot \mathbf{a}_1)} \sum_{n_2=1}^{N_2} e^{2\pi i n_2(\mathbf{h} \cdot \mathbf{a}_2)} \sum_{n_3=1}^{N_3} e^{2\pi i n_3(\mathbf{h} \cdot \mathbf{a}_3)}. \quad 2.4. 24$$

Sendo a soma de uma série geométrica dada por:

$$\sum_{j=0}^N r^j = \left[\frac{r^{N+1} - 1}{r - 1} \right] \quad 2.4. 25$$

renumeramos as células unitárias de $n_j = 0$ a $n_j = N_j - 1$, e multiplicando e dividindo o termo de cada somatório de \mathcal{S} por $e^{-\pi i(\mathbf{h} \cdot \mathbf{a}_j)}$, temos que:

$$\mathcal{S} = e^{i\Phi} \left[\frac{\sin(\pi N_1 \mathbf{h} \cdot \mathbf{a}_1)}{\sin(\pi \mathbf{h} \cdot \mathbf{a}_1)} \right] \left[\frac{\sin(\pi N_2 \mathbf{h} \cdot \mathbf{a}_2)}{\sin(\pi \mathbf{h} \cdot \mathbf{a}_2)} \right] \left[\frac{\sin(\pi N_3 \mathbf{h} \cdot \mathbf{a}_3)}{\sin(\pi \mathbf{h} \cdot \mathbf{a}_3)} \right] \quad 2.4. 26$$

onde $\Phi = [\pi \mathbf{h} \cdot \mathbf{a}_1(1 - N_{a1})] + [\pi \mathbf{h} \cdot \mathbf{a}_1(1 - N_{a2})] + [\pi \mathbf{h} \cdot \mathbf{a}_1(1 - N_{a3})]$. Vemos que \mathcal{S} só é diferente de zero quando:

$$\mathbf{h} \cdot \mathbf{a}_i = h_i; \quad 2.4. 27$$

que é a mesma condição dada na equação 2.4. 14. Ou seja, após atravessar um cristal real finito, a radiação espalhada apresentará interferências construtivas apenas em pontos do espaço recíproco definido pela equação 2.4. 11.

Agora voltamos para a análise da amplitude total espalhada pelo cristal. Lembramos que o espalhamento acontece principalmente por elétrons e que a amplitude espalhada depende também do poder de espalhamento do espalhador. Sendo ε o poder de espalhamento de um elétron, podemos reescrever 2.4. 19 adicionando-o como constante multiplicativa:

$$A_T \propto \varepsilon \int \rho_{crist}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} = \varepsilon \cdot \mathcal{F}_r[\rho_{crist}(\mathbf{r})](\mathbf{h}) \quad 2.4. 28$$

Das equações 2.4. 22 e 2.4. 23:

$$A_T \propto \varepsilon \int_{cel} \rho_{cel}(\mathbf{r}) e^{2\pi i(\mathbf{h} \cdot \mathbf{r})} d\mathbf{r} \quad 2.4. 29$$

Numa primeira aproximação, consideramos que a distribuição de cargas em torno dos átomos é esfericamente simétrica e centrada no núcleo de cada átomo j cuja posição é especificada pelo vetor \mathbf{r}_j . Reescrevemos $\rho_{cel}(\mathbf{r})$ em função de uma nova variável \mathbf{r}' e da densidade ρ_{ej} de espalhadores no átomo j , tal que:

$$\mathbf{r} = \mathbf{r}_j + \mathbf{r}' \rightarrow d\mathbf{r} = d\mathbf{r}' \quad 2.4. 30$$

$$\rho_{ej}(\mathbf{r}') = \begin{cases} \rho_{ej}(\mathbf{r}') & \text{se } r' \leq \text{raio atômico } R \\ 0 & \text{se } r' \geq \text{raio atômico } R \end{cases} \quad 2.4. 31$$

e portanto:

$$\rho_{cel}(\mathbf{r}') = \sum_{j=1}^N \rho_{ej}(\mathbf{r}') \quad 2.4. 32$$

onde N é o número de átomos na célula unitária. Reescrevemos 2.4.26:

$$A_T \propto \sum_{j=1}^N \left[\varepsilon \int_0^R \rho_{ej}(\mathbf{r}') e^{2\pi i \mathbf{h} \cdot \mathbf{r}'} d\mathbf{r}' \right] e^{2\pi i \mathbf{h} \cdot \mathbf{r}_j} \quad 2.4. 33$$

A soma é conhecida como Fator de Estrutura, $F(\mathbf{h})$. Sua forma mais conhecida é:

$$F(\mathbf{h}) = \sum_{j=1}^N f_{ej} e^{2\pi i (\mathbf{h} \cdot \mathbf{r}_j)} \quad 2.4. 34$$

onde

$$f_{ej} = \varepsilon \int_0^R \rho_{ej}(\mathbf{r}') e^{2\pi i (\mathbf{h} \cdot \mathbf{r}')} d\mathbf{r}' \quad 2.4. 35$$

é conhecido como fator de forma atômico, e guarda informação sobre a distribuição de cargas no átomo j . É importante enfatizar que o Fator de Estrutura é uma grandeza complexa, e pode ser escrito em termos de seu módulo, F_{hkl} , e sua parte complexa com fase φ_{hkl} :

$$F(\mathbf{h}) = F_{hkl} e^{i\varphi_{hkl}} \quad 2.4. 36$$

A densidade eletrônica da célula unitária do cristal, que é a informação estrutural buscada nos experimentos de difração, é a transformada inversa de Fourier do fator de estrutura, dada por:

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_{hkl} F(\mathbf{h}) e^{-2\pi i \mathbf{h} \cdot \mathbf{r}} \quad 2.4. 37$$

onde V é o volume da célula unitária.

Vemos, portanto, que a densidade eletrônica dentro de um cristal (posição dos espalhadores no cristal) está relacionada com a posição dos picos de difração por uma transformada de Fourier. Isso nos permite utilizar experimentos de difração de raios X (XRD) por um cristal na elucidação da estrutura cristalina. O experimento de XRD será detalhado nas seção 2.5.

Fator de deslocamento isotrópico e anisotrópico

Até agora, tratamos o problema da difração por um cristal assinalando ao núcleo de cada átomo uma posição determinada por um vetor \mathbf{r}_j . Entretanto, é necessário considerar que a energia térmica do ambiente faz com que os átomos oscilem em torno de suas posições de equilíbrio. Analogamente, devido a defeitos intrínsecos do processo de crescimento dos cristais, podemos supor que alguns átomos possam não estar localizados exatamente na posição de equilíbrio esperada em um dada célula unitária. Para levar esses efeitos em consideração nos cálculos apresentados na última seção, é necessário reescrever a função densidade eletrônica de cada átomo, que agora será dada pela convolução:

$$\rho_v(\mathbf{r}) = \int_V \rho_{ej}(\mathbf{r} - \mathbf{u}) p(\mathbf{u}) d\mathbf{u} \quad 2.4. 38$$

onde $\rho_v(\mathbf{r})$ é a densidade eletrônica para o átomo fora da posição de equilíbrio, $\rho_{ej}(\mathbf{r} - \mathbf{u})$ é a densidade eletrônica no ponto \mathbf{r} quando o átomo j estiver centrado num ponto definido pelo vetor \mathbf{u} e $p(\mathbf{u})$ é a probabilidade desse átomo estar centrado em \mathbf{u} . Pelo teorema da convolução, a transformada de Fourier de $\rho_v(\mathbf{r})$ será:

$$f_v(\mathbf{h}) = f_{ej}(\mathbf{h})q(\mathbf{h}) \quad 2.4. 39$$

onde $f_{ej}(\mathbf{h})$ e $q(\mathbf{h})$ são as transformadas de Fourier de $\rho_v(\mathbf{r})$ e $p(\mathbf{u})$, respectivamente. O fator $q(\mathbf{h})$ é conhecido como fator de Debye-Waller.

Se considerarmos inicialmente que os átomos vibram em torno de suas posições de equilíbrio de forma isotrópica, ou seja, se movimentam de forma esfericamente simétrica, podemos aproximar a $p(\mathbf{u})$ por uma Gaussiana:

$$p(\mathbf{u}) \cong \frac{1}{\sqrt{2\pi U}} e^{-u^2/2U} \quad 2.4. 40$$

onde $U = \langle \mathbf{u}^2 \rangle$ é o deslocamento quadrático médio da posição de equilíbrio do referido átomo. Dessa forma:

$$q(\mathbf{h}) = e^{(-2\pi^2 U h^2)} = e^{(-B \sin^2 \theta / \lambda^2)} \quad 2.4. 41$$

onde:

$$B = 8\pi^2 U. \quad 2.4. 42$$

Utilizando as equações 2.4. 35 e 2.4. 41, escrevemos o fator de forma para um átomo que vibra de forma isotrópica:

$$f_{ej} = \varepsilon. e^{(-B \sin^2 \theta / \lambda^2)} \int_0^R \rho_{ej}(\mathbf{r}') e^{2\pi i(\mathbf{h} \cdot \mathbf{r}')} d\mathbf{r}'. \quad 2.4. 43$$

Sabemos, entretanto que, no geral, um átomo na estrutura cristalina não tem liberdade para vibrar igualmente em todas as direções. De fato, as superfícies de probabilidade serão elipsoides centrados na posição média do átomo. No espaço recíproco, a quantidade U é agora um tensor $U = \langle \mathbf{u} \cdot \mathbf{u} \rangle$, portanto o fator de correção $q(\mathbf{h})$ se torna:

$$q(\mathbf{h}) = \exp(-2\pi^2 \langle \mathbf{h} | U | \mathbf{h} \rangle) \quad 2.4.44$$

onde:

$$\langle \mathbf{h} | U | \mathbf{h} \rangle = U_{11}h^2 + U_{22}k^2 + U_{33}l^2 + U_{23}kl + U_{13}hl + U_{12}hk \quad 2.4.45$$

onde $U_{ij} = U_{ji}$ são as componentes do tensor.

Histórico

Os raios X foram descobertos e descritos em 1895 por Wilhelm Röntgen e foram imediatamente utilizados em hospitais para radiografias. Seu comprimento de onda, que vai de aproximadamente 0.1 a 100Å, foi primeiramente medido por Sommerfeld em 1912. Posteriormente, Max von Laue teorizou a difração dos raios X por uma estrutura periódica, o que foi comprovado experimentalmente com o auxílio dos alunos de Röntgen, Friederich e Knipping. Em 1913, William Lawrence Bragg, analisando os trabalhos de Laue, interpretou que cada plano de átomos num cristal funcionava como uma superfície refletora de raios X, o que levava à interferência dos raios refletidos e portanto explicava os picos de difração observados. Suas ideias traduzidas para forma matemática levavam à equação hoje chamada Lei de Bragg mostrada na equação 2.4. 15.

O pai de W. L. Bragg, Willian Henry Bragg, então projetou e construiu o primeiro espectrômetro de raios X, que junto às ideias de seu filho, permitiu que inúmeras estruturas de materiais cristalinos fossem determinadas. Foi na década de 70 que o advento do uso de radiação sincrotron na produção de raios X possibilitou um avanço ainda maior no estudo com esse tipo de radiação e conseqüentemente mais avanços nos estudos que envolvem sua aplicação.

2.5 Experimento de difração de raios X por monocristal

Setup experimental e coleta de dados

A natureza da interação dos raios X com estruturas com organização de longo alcance torna a difração de raios X por cristais uma poderosa técnica para determinação estrutural. Isso porquê, como mostrado no capítulo anterior, a densidade eletrônica em um cristal se relaciona matematicamente com a amplitude da onda difratada por ele e a intensidade dos máximos de interferência construtiva observados se relaciona com a posição dos espalhadores dentro da célula unitária do cristal.

O aparato utilizado no experimento de difração de raios X está ilustrado na Figura 2.5.1 e consiste essencialmente de uma fonte de radiação X, um porta amostra acoplado a um goniômetro que permite mudar a orientação da amostra em relação ao feixe, e um detector de raios X. As fontes de raios X mais comumente utilizadas hoje são tubos de raios X, em aparatos mais compactos, e fontes de luz síncrotron.

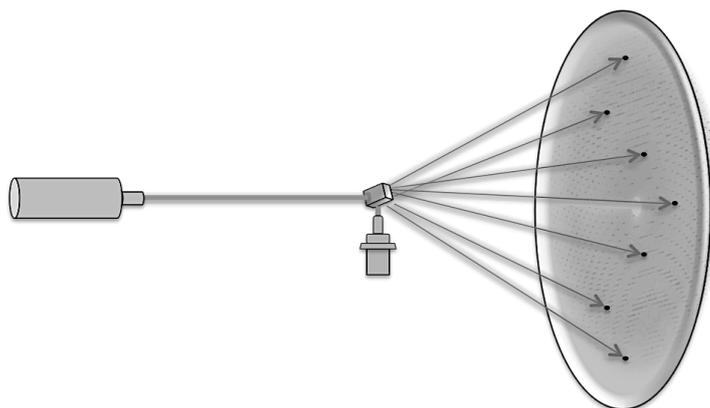


Figura 2.5.1 – Diagrama esquemático do aparato utilizado no experimento de difração de raios X.

O experimento então consiste em coletar um conjunto de imagens do padrão de difração do cristal para diferentes orientações da amostra com relação ao feixe. Em cada imagem são extraídas as intensidades dos máximos de interferência

construtiva e suas posições relativas a um sistema de referência arbitrário. É importante ressaltar que o detector de raios X possui tamanho limitado e, em grande parte dos casos, é uma superfície plana. As imagens são, portanto, projeções bidimensionais de partes do espaço recíproco.

Processamento de dados

Uma vez de posse das imagens do padrão de difração do cristal, um conjunto de algoritmos computacionais é utilizado no processamento e interpretação das imagens para construção do espaço recíproco tridimensional.

O primeiro passo no processamento de dados é a determinação da célula unitária e da rede de Bravais do cristal. Esse processo é chamado indexação e está esquematizado no fluxograma da Figura 2.5.2 [50]. Inicialmente, uma pequena parte dos dados de difração coletados que engloba partes distintas do espaço recíproco é utilizada na tentativa de determinar os vetores que melhor descrevem a célula unitária primitiva do cristal. Lembramos que os vetores \mathbf{a}_1^* , \mathbf{a}_2^* e \mathbf{a}_3^* que descrevem corretamente a célula unitária formam uma base para o espaço recíproco tal que a posição \mathbf{h} de todas as reflexões medidas podem ser descritas como uma combinação linear deles:

$$\mathbf{h} = h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^* \quad 2.5.1$$

onde h , k e l são números inteiros, chamados de índices de Miller.

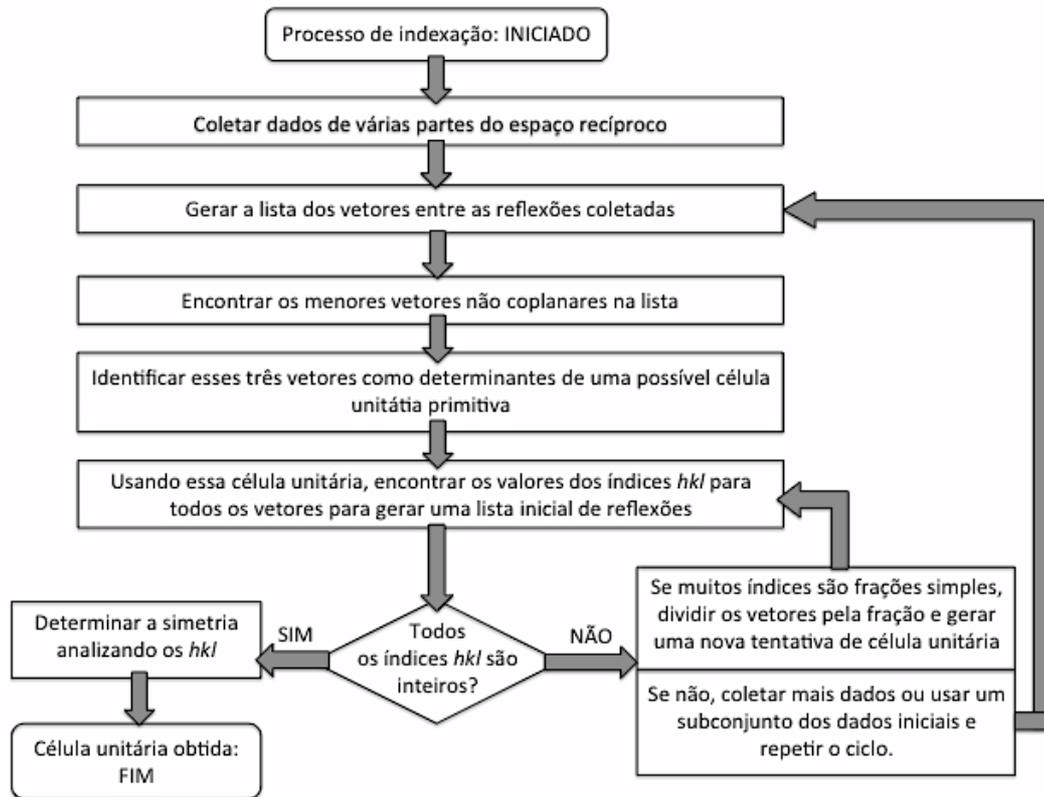


Figura 2.5.2 – Fluxograma do processo de indexação dos dados de difração de raios X

O segundo passo é a determinação da posição da célula unitária com relação ao instrumento de medida uma vez que a posição do cristal em relação ao feixe de raios X varia ao longo dos experimentos de difração. É necessário relacionar os eixos do referencial fixo (referencial do equipamento) com os eixos que definem a célula unitária no espaço recíproco e conseqüentemente com os eixos que a definem no espaço real.

Com base na célula unitária determinada e uma vez assinaladas as orientações corretas da amostra a cada imagem, é possível integrar os dados, atribuindo para cada reflexão uma intensidade $I(\mathbf{h})$.

Problema das fases

O objetivo do experimento de difração de raios X é elucidar a densidade eletrônica $\rho(\mathbf{r})$ dentro do cristal a partir das informações guardadas na

amplitude da onda difratada, como vimos na seção 2.4. A quantidade que pretendemos determinar é o fator de estrutura (equação 2.4. 34), cuja transformada de Fourier nos daria $\rho(\mathbf{r})$. Entretanto, medimos apenas a intensidade da onda difratada, $I(\mathbf{h})$, que é proporcional ao quadrado da amplitude e consequentemente proporcional ao quadrado do Fator de Estrutura:

$$I(\mathbf{h}) \propto |F(\mathbf{h})|^2. \quad 2.5.2$$

Sendo o Fator de estrutura uma grandeza complexa, perdemos portanto a informação contida nas fases quando medimos apenas a intensidade. A incoerência temporal e espacial das fontes de raios X mais comumente utilizadas não permite medir a relação final entre as fases das ondas difratadas. Por conseguinte, para elucidar a densidade eletrônica dentro do cristal, uma série de métodos, nomeados métodos de faseamento, são utilizados para deduzir, medir indiretamente ou especular a relação entre as fases dos módulos dos fatores de estrutura medidos. Um método comumente utilizado para resolver o problema da perda da informação das fases de medidas de difração de raios X por cristais de proteínas é a Substituição Molecular, que será detalhada na próxima seção.

2.6 Solução e refinamento da estrutura

A função de Patterson

Como descrito na seção anterior, construir a densidade eletrônica requer que conheçamos a magnitude e a fase de cada fator de estrutura $F(\mathbf{h})$ associado a cada reflexão, mas o experimento não permite medir a relação de fases. É possível, no entanto, determinar os valores de $|F(\mathbf{h})|^2$ aplicando os devidos fatores de correção à quantidade medida para obtermos as informações estruturais desejadas. Para tal será necessário definir a função de Patterson:

$$P(\mathbf{u}) = \frac{1}{V} \sum_h \sum_k \sum_l |F(\mathbf{h})|^2 e^{-2\pi i(hu+kv+lw)} = \frac{1}{V} \mathcal{F}^{-1}(|F(\mathbf{h})|^2) \quad 2.6.1$$

onde $\mathbf{u} = \mathbf{u}(u, v, w)$ é um vetor com as mesmas dimensões dos vetores do espaço real \mathbf{r} . Vemos que $P(u, v, w)$ é a transformada inversa de Fourier de $|F(\mathbf{h})|^2$ e, portanto, pode ser calculada diretamente dos dados experimentais. Para relacionar a função de Patterson com a função densidade eletrônica, escrevemos:

$$|F(\mathbf{h})|^2 = F(\mathbf{h})F(\mathbf{h})^* \quad 2.6.2$$

onde $F(\mathbf{h})^*$ é o complexo conjugado de $F(\mathbf{h})$, de tal forma que a função de Patterson pode ser reescrita como:

$$P(u, v, w) = \frac{1}{V} \mathcal{F}^{-1}(F(\mathbf{h})F(\mathbf{h})^*) = \frac{1}{V} [\mathcal{F}^{-1}(F(\mathbf{h})) * \mathcal{F}^{-1}(F^*(\mathbf{h}))]. \quad 2.6.3$$

Da seção 2.4, vimos que:

$$\rho(x, y, z) = \frac{1}{V} \mathcal{F}^{-1}(F(\mathbf{h})) \quad 2.6.4$$

e utilizando uma mudança de variáveis é possível mostrar que⁸:

$$\rho(-x, -y, -z) = \frac{1}{V} \mathcal{F}^{-1}(F^*(\mathbf{h})) \quad 2.6.5$$

de tal forma que:

⁸ Para demonstração, ver [1] p.435

$$P(\mathbf{u}) = V[\rho(\mathbf{r}) * \rho(-\mathbf{r})], \quad 2.6.6$$

ou seja,

$$P(\mathbf{u}) = V \int \rho(\mathbf{r}) * \rho(\mathbf{r} + \mathbf{u}) d\mathbf{r}, \quad 2.6.7$$

A função de Patterson pode então ser interpretada como a convolução entre a função densidade eletrônica com sua imagem centrossimétrica, multiplicada pelo volume da célula unitária[18].

É importante destacar algumas das propriedades da função de Patterson:

- é uma função centrossimétrica;
- contém $N(N - 1)$ picos fora da origem, onde N é o número de picos da função de densidade eletrônica, que acontecem para vetores \mathbf{u} correspondentes aos vetores interatômicos na célula unitária;
- picos mais próximos tem \mathbf{u} correspondentes à vetores intramoleculares, enquanto picos mais afastados da origem tem \mathbf{u} correspondentes à vetores intermoleculares.

Substituição molecular

O método de substituição molecular (MR) é comumente utilizado para resolver o problema das fases em casos específicos onde a molécula com estrutura a ser determinada (alvo) é similar a outra cuja estrutura é conhecida (modelo de busca).

Dispõe-se, inicialmente, de um conjunto de dados de difração de raios X e dos dados da estrutura tridimensional do modelo de busca. Essencialmente, o método de MR consiste em encontrar a orientação e a posição do modelo de

busca na célula unitária que geraria um conjunto de fatores de estrutura calculados ($F_{calc}(\mathbf{h})$) com valores próximos aos valores medidos ($F_{obs}(\mathbf{h})$).

Começamos representando as coordenadas dos átomos da molécula do modelo de busca por uma matriz da forma:

$$\mathbf{x} = \begin{pmatrix} x_1 & x_2 & x_3 & x_j \\ y_1 & y_2 & y_3 & y_j \\ z_1 & z_2 & z_3 & z_j \end{pmatrix} \quad 2.6.8$$

e precisamos encontrar a matriz de rotação $[C]$ e o vetor translação \mathbf{d} que a superpõe com a molécula alvo, cujos átomos tem coordenadas representadas pela matriz \mathbf{x}' :

$$\mathbf{x}' = [C]\mathbf{x} + \mathbf{d} \quad 2.6.9$$

onde

$$[C] = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \quad \text{e} \quad \mathbf{d} = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}. \quad 2.6.10$$

Em suma, o problema consiste em determinar os 3 ângulos de rotação e as 3 coordenadas do vetor translação que minimizam a diferença entre $F_{obs}(\mathbf{h})$ e $F_{calc}(\mathbf{h})$.

Uma possível estratégia seria avaliar todas as possíveis posições da molécula de modelo de busca na célula unitária do cristal. Mas proteínas são moléculas grandes e relativamente compactas, cujos cristais costumam apresentar grandes canais de solvente entre as moléculas. Dessa forma, as distâncias intermoleculares são majoritariamente maiores que as intramoleculares e é possível distinguir os respectivos vetores nos mapas de Patterson. Assim esse processo fica bem mais simples quando dividido em duas etapas: encontrar a função de rotação e encontrar a função translação.

O objetivo da função rotação é determinar a orientação da molécula modelo na célula unitária do cristal. Para tal avaliamos a concordância entre a função $P(u, v, w)$ para os dados de difração medidos, P_A , e para a molécula modelo girada, $\Re P_M$, utilizando a função proposta por Rossmann e Blow [51]:

$$R = \int P_A(\mathbf{u}) \Re P_M(\mathbf{u}) d\mathbf{u} \quad 2.6. 11$$

que assume valor máximo quando as duas moléculas se superpõe melhor. Para avalia-la, é necessário definir o passo da procura da variável angular, os limites de resolução dos dados medidos, a molécula modelo e o raio de integração. O último deve ser tal que englobe os vetores intramoleculares, invariantes a translações, e exclua os vetores intermoleculares.

A função translação, por sua vez, objetiva posicionar corretamente a molécula de busca com relação aos elementos de simetria do cristal e os métodos para sua busca são um pouco mais variados e sofisticados. Os métodos baseados na função de Patterson, particularmente, assumem que apenas quando todas as moléculas na célula unitária estiverem posicionadas corretamente, os vetores intermoleculares serão compatíveis com o mapa de Patterson calculado para os dados medidos. Novamente, comparamos mapas de Patterson, dessa vez utilizando os picos correspondentes aos vetores intermoleculares. Podemos definir a função de translação matematicamente por:

$$T(\mathbf{t}) = \int P_M(\mathbf{u}, \mathbf{t}) P_A(\mathbf{u}) d\mathbf{u} \quad 2.6. 12$$

onde \mathbf{t} é um vetor que define a origem com relação a um elemento de simetria. Uma abordagem mais detalhada das funções de translação pode ser encontrado nas referências [48]–[51].

Tendo em mãos a orientação do modelo de busca, as intensidades medidas e as fases dos fatores de estrutura calculados para o modelo de busca, podemos gerar

um mapa de densidade eletrônica experimental $\rho_{obs}(\mathbf{r})$ inicial. Esse mapa é obtido pelo cálculo da transformada de

$$\rho_{obs}(\mathbf{r}) = \frac{1}{V} \mathcal{F}^{-1}(F_{obs}(\mathbf{h})) = \frac{1}{V} \mathcal{F}^{-1}(\sqrt{I_{obs}(\mathbf{h})} e^{i\varphi_{modelo}}) \quad 2.6.13$$

onde $I_{obs}(h)$ são as intensidades medidas e φ_{modelo} são as fases advindas do modelo de busca rotacionado e translacionado de acordo com 2.6.9. Utilizando a estrutura primária da proteína, é então possível construir um modelo estrutural que se ajuste aproximadamente à densidade eletrônica experimental, e então um fator de estrutura (F_{calc}) é calculado para esse modelo. As fases iniciais do modelo obtido por MR (φ_{modelo}) são então substituídas pelas fases calculadas (φ_{calc}) e um novo mapa de densidade eletrônica experimental é obtido $\rho_{obs}(\mathbf{r}) = \frac{1}{V} \mathcal{F}^{-1}(F_{obs}(h)) = \frac{1}{V} \mathcal{F}^{-1}(\sqrt{I_{obs}(h)} e^{i\varphi_{calc}}$. Resumidamente, cada ajuste/melhoria no modelo estrutural da proteína permite o cálculo de um novo mapa experimental que por sua vez leva a uma melhoria do modelo. Este processo é continuado até que o valor da diferença $F_{obs} - F_{calc}$ seja a menor possível.

Apesar de sua limitação inicial, o número de estruturas depositadas em bancos de dados, como o *Protein Data Bank* (PDB)[3], faz com que o método de MR seja responsável pela solução (determinação estrutural) de aproximadamente 70% das macromoléculas depositadas [2], [3]. Para que o MR seja efetivo, é necessário que as duas estruturas apresentem uma similaridade mínima de 25% na sequência de aminoácidos que compõe suas estruturas primárias [57], mas o objetivo ao buscar estruturas primárias similares é encontrar estruturas tridimensionais parecidas, de forma que, em muitos casos, uma proteína da mesma família, ainda que com similaridade menor que 25%, é um modelo melhor para ser utilizado na substituição molecular.

Softwares e indicadores de qualidade

Todo o trabalho com os dados de difração de raios X conta com diversos softwares, dedicados a diferentes tarefas e com diferentes metodologias. Em todos eles, uma série de parâmetros são utilizados para indicar a qualidade dos dados de difração de raios X, do processamento e do ajuste do modelo proposto ao mapa de densidade eletrônica experimental.

Nesse trabalho, os processamentos de dados de XRD contaram com os softwares: *HKL2000* [58] e *XDS* [59], [60]. Ambos são programas comumente utilizados no processamento de dados de XRD e apresentam os métodos de indexação mais robustos dentre os softwares desenvolvidos para o mesmo propósito. Uma comparação entre os programas de processamento de dados de XRD foi feita e reportada por Gilski, 2012 [61].

Os parâmetros avaliados nas estatísticas dos processamentos de dados estão listados a seguir junto à uma breve explicação de seus significados.

- Resolução: é o espaçamento d entre os planos da rede cristalina (equação 2.4. 15). Medida em angstroms (Å), a resolução define o nível de detalhe, ou seja, a menor distância entre partes da estrutura que podem ser resolvidas no mapa de densidade eletrônica.
- Reflexões: número total de picos de interferência construtiva indexadas pelo software durante o processamento de dados.
- Reflexões únicas: número de reflexões não equivalentes por simetria indexadas pelo software durante o processamento de dados.
- R_{merge} : é um parâmetro que mede o desvio dos valores das intensidades medidas com relação ao valor médio calculado. Lembramos que muitas reflexões são medidas mais de uma vez, junto às suas equivalentes por simetria, de tal forma que intensidade assinalada a elas são a média desses valores. Por definição,

$$R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i \langle I_i(hkl) \rangle} \quad 2.6. 14$$

onde $h, k, e l$ são os índices de Miller e i enumera reflexões equivalentes.

- R_{meas} : assim como R_{merge} , é utilizado na inferência da exatidão do valor médio calculado para as intensidade de reflexões equivalente. Entretanto, esse é ponderado de forma à ser independente da multiplicidade (número de reflexões equivalente) observada. Por definição,

$$R_{meas} = \frac{\sum_{hkl} \left(\frac{N_{hkl}}{N_{hkl} - 1} \right)^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i \langle I_i(hkl) \rangle} \quad 2.6. 15$$

onde N_{hkl} é o número de vezes que a reflexão hkl foi medida.

- $\langle I/\sigma(I) \rangle$: é o valor médio da razão entre sinal e ruído, onde I é a intensidade de cada reflexão e $\sigma(I)$ o erro intrínseco da intensidade estimado para cada reflexão .
- Completeza: parâmetro que indica a porcentagem de reflexões únicas medidas em relação ao número de reflexões únicas previstas teoricamente.

Após o processamento de dados, contamos com a plataforma *CCP4* [62], onde estão reunidos vários softwares para cristalografia de macromoléculas por difração de raios X, utilizados nas etapas do trabalho de determinação, refinamento e validação estrutural. As substituições moleculares foram executadas utilizando o programa *PHASER* [63], o alinhamento de cadeias proteicas baseado na estrutura delas contou com o programa *GESAMT* [64], o refinamento dos modelos propostos com o software *REFMAC5* [65], a inspeção visual e a manipulação dos modelos com o *COOT* [66], [67] e as figuras com o software *CHIMERA* [68].

Os parâmetros avaliados nas estatísticas dos refinamentos dos modelos propostos estão listados a seguir junto à uma breve explicação de seus significados.

- R_{fator} : mede a discrepância entre as amplitudes dos fatores de estrutura observados, F_{obs} , e as amplitudes dos fatores de estrutura calculados para o modelo proposto. Por definição,

$$R_{fator} = \frac{\sum |F_{obs} - F_{calc}|}{\sum F_{obs}} . \quad 2.6.16$$

- R_{free} : calculado de modo análogo ao R_{fator} , entretanto apenas para um subconjunto de reflexões medidas que não são levadas em consideração em nenhuma etapa do refinamento do modelo.
- RMS ângulos/ligações: indicam o quanto os valores dos parâmetros geométricos (ângulos e comprimento de ligações) no modelo proposto diferem dos valores considerados típicos.

Uma discussão completa sobre os parâmetros considerados no processamento de dados de difração de raios X por monocristal de proteína foi feita por Wlodawer *et al.* [69]. Mais detalhes também podem ser encontrados nas referências: [18], [70].

Esfingomielinase D

3.1 Introdução

As fosfolipases (PL) são enzimas que hidrolisam fosfolipídios, como fosfatidilcolina, fosfatidilinositol, fosfatidilserina e esfingomielina, e também outras moléculas que compõem a membrana celular. Ao analisar a estrutura básica de um glicerofosfolipídeo mostrada na Figura 3.1.1, vemos que existem quatro possíveis ligações éster que podem ser alvo das fosfolipases. De acordo com a ligação que hidrolisam elas são separadas em 4 classes: fosfolipase A1, fosfolipase A2, fosfolipase C e fosfolipase D[71].

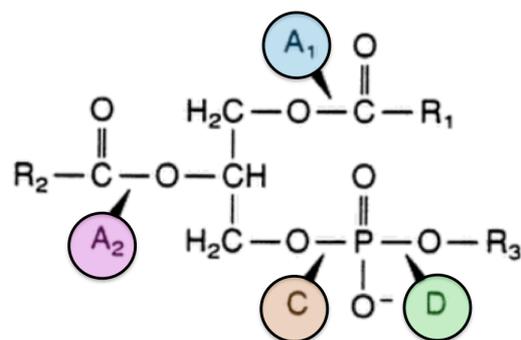


Figura 3.1.1 – Estrutura geral de um fosfolipídeo e os quatro sítios, A₁, A₂, C e D, que podem sofrer hidrólise pelas fosfolipases. (Imagem adaptada de [71])

As fosfolipases que clivam a esfingomielina em fosforilcolina e ceramida são especificamente denominadas esfingomielinases (SMases) e atuam na hidrólise do fosfolipídeo através de um mecanismo similar aos das fosfolipases C. Elas vem sendo estudadas nas últimas décadas desde a descoberta da via de transdução de sinal da esfingomielina [72]–[75], mas também pela participação da ceramida em diversos processos celulares como apoptose, diferenciação e proliferação, o

que motiva o estudo das SMases, que são grandes fontes desse lipídeo e estão muitas vezes associadas ao processo de morte celular [76], [77].

A esfingomielinase D

As esfingomielinases D (SMaseD) são um tipo específico de SMase que cliva o fosfolipídeo no sítio D, atividade rara para esse tipo de enzimas. A atividade da SMase D é bem conhecida nas bactérias *Corynebacterium pseudotuberculosis*, *Corynebacterium ulcerans*, e, no veneno das aranhas do gênero *Loxocles* (no Brasil, a aranha-marrom é seu principal representante) [78]. Contudo, estudos mais recentes permitiram a identificação das SMases D em algumas espécies de fungos e em outros organismos patogênicos [79].

As bactérias do gênero *Corynebacterium* estão envolvidas em infecções de inúmeras populações de diversos animais. Particularmente, *C. Ulcerans* está principalmente envolvida em infecções em bovinos e equinos, sendo ela a causadora da mastite bovina, mas também é capaz de infectar humanos [80], [81]. As SMases D são um determinante essencial da virulência dessa bactéria, contribuindo para sua propagação do local inicial da infecção para sítios secundários dentro do hospedeiro [79].

A SMases D identificadas por análises sequenciais em vários organismos patogênicos pertencentes a gêneros distintos aparentemente compartilham um sítio ativo conservado e um motivo C-terminal. Contudo, poucas estruturas tridimensionais dessa família de proteínas foram determinadas e depositadas no PDB.

A SMase D de *C. ulcerans* possui aproximadamente 31,3kDa⁹ de peso molecular, estimado a partir dos 284 aminoácidos da estrutura primária mostrada na sequência a seguir:

**MTPVSHDAASTGNRPVYAIHRVLTQTQGVDDAVAIGANALEIDFTAWRGGWWADHDGIPTS
AGATAEAIFKHIAEKRRQGANITFTWLDIKNPDYCTDPDSVCSINALRDLARKYLEPAGVRVL**

⁹ 1 Da = 1 u.m.a = 1/12 da massa de um átomo de carbono-12

YGFYKTVGGPGWKTITSDLRDKEAVALSGPTRDVLNDFAKAGDKILTKQKIADYGYDINQG
FGNCYGDGNKTC DQLRKSSEARDQGQLGKTFGWTTTGQDDRVDLLGKAHVDGMIFGFK
VTHFYRHADTENSFKAIKTWVDKHS DTHHLATAADNPW

A SMase D de *C. ulcerans* apresenta baixa identidade de sequência (< 20%) com outras SMases e PLDs depositadas (PDB id's: 3R1H [82], 4RW3 [83], 2F9R e 1XX1 [84], 4Q6X [85]).

Nesse trabalho, apresentamos estudos estruturais e a cristalização da SMase D de *C. ulcerans*. Novas informações estruturais desse grupo de enzimas são de grande valor na compreensão do papel e da evolução das SMases D, principalmente dadas as suas propriedades patogênicas.

3.2 Procedimentos e discussão

Expressão em *Escherichia coli*

O vetor de expressão pET28a-TEV contendo a sequência codificante de uma SMase D de *C. ulcerans* com uma cauda de histidinas (*His-tag*) no N-terminal que foi utilizado nos experimentos com *E. coli* foi cedido pela prof. Dra. Liza Felicori do Departamento de Bioquímica e imunologia do ICB da UFMG. Os protocolos de expressão e purificação da enzima já haviam sido estabelecidos pelo grupo de pesquisa da Dra. Felicori, e foram otimizados nesse trabalho.

Células de *E. coli* BL21 foram devidamente transformadas para receber o vetor de expressão pET28a-TEV e devidamente plaqueadas. Colônias únicas foram inoculadas separadamente em meio de cultura 2xYT¹⁰ esterilizado contendo os antibióticos gentamicina e canamicina, ou, alternativamente, fez-se o pré-inóculo a partir de alíquotas de culturas conservadas em glicerol. Essa alíquota inicial de meio de cultura que recebeu as colônias foi então mantida a 37°C e sob agitação de 200rpm, para garantir a aeração do meio, por 16h, tempo suficiente para crescimento significativo da população de células. Em seguida, as culturas iniciais foram diluídas na proporção 1:50 em meio 2xYT esterilizado sem

¹⁰ Composição: triptona 16 g/L; extrato de levedura 10g/L; cloreto de sódio 5 g/L

antibiótico e cultivadas por 3h, novamente a 37°C e 200rpm. Quando o crescimento das células chegou aproximadamente à metade da fase de crescimento logarítmico [86], em média três horas após o inóculo, fez-se a indução da expressão proteica com IPTG a uma concentração final de 0.5mM. A fase de crescimento da cultura de células foi monitorada através da medida da densidade óptica do meio [87] utilizando o espectrofotômetro *Biochrom WPA CO8000*. A expressão prosseguiu por mais 4 horas, de novo a 37°C e 200rpm. Alíquotas da cultura no início e no término do processo de expressão foram coletadas e submetidas à eletroforese em gel de poliacrilamida e dodecil-sulfato de sódio (SDS-PAGE [88]) (Figura 3.2.3).

Após o período de expressão, o meio com as células foi centrifugado a 5000g e a 4°C por 15 minutos, forçando a precipitação das células. O sobrenadante foi descartado e o precipitado foi congelado (-20°C) ou imediatamente submetido a lise.

Lise celular

O precipitado celular proveniente de 1,5 litros de cultura de células obtido na etapa de expressão foi ressuscitado em 50mL de tampão A (Tris-HCl 50mM pH 7,45; NaCl 300mM, MgSO₄ 5mM, PMSF 1mM, Imidazol 20mM) e lisado sob pressão de 15000 a 20000psi em homogeneizador *Emulsiflex®- C3* (Avestin). Para garantir a lise de todas as células, fez-se em média 3 passagens pelo aparelho. Com intuito de separar o sobrenadante da fração insolúvel o lisado foi então centrifugado a 10000g e a 4°C por 20 minutos.

Purificação da proteína recombinante

A obtenção da proteína recombinante pura a partir da fração solúvel do produto da lise bacteriana foi feita em etapas sequenciais. A etapa 1 consistiu em uma cromatografia de afinidade a níquel (AF) em sistema *ÄKTAprime plus* (GE Healthcare) cujo cromatograma de absorção de luz UV (curva azul) é mostrado na Figura 3.2.1. Em coluna *HisTrap HP* de 5mL (GE Healthcare) previamente

ambientada com tampão A, aplicou-se o extrato da lise e em seguida a coluna foi lavada com tampão A. Observa-se uma alta absorbância inicial devido às demais proteínas e componentes do extrato celular bruto. A lavagem prosseguiu até que nenhuma leitura de absorbância de UV fosse observada, garantindo que as demais proteínas bacterianas que não possuíssem afinidade pela matriz de níquel fossem removidas da coluna. A eluição das proteínas aderidas à coluna foi feita em concentração crescente de imidazol utilizando tampão B (Tris-HCl 50mM pH 7,45; NaCl 300mM, MgSO₄ 5mM, PMSF 1mM, Imidazol 500mM), como pode ser visto na curva cinza da Figura 3.2.1. À medida que a concentração de imidazol é aumentada, observa-se dois picos de alta absorbância, que acredita-se serem referentes à SMase D. O eluente da coluna foi recolhido em frações de 3mL de tal forma que foi possível separar as que correspondiam a cada pico e à superposição de ambos. As amostras foram colocadas em gelo imediatamente após sua coleta no equipamento. As frações correspondentes ao primeiro pico apresentaram uma instabilidade formando aglomerados turvos ainda que mantidas em baixa temperatura. Já as correspondentes ao segundo pico mantiveram o aspecto transparente e translúcido e apenas elas foram submetidas à segunda etapa de purificação. Acreditamos que o primeiro pico também corresponda a SMase D, entretanto em algum estado de agregação que não favorece sua estabilidade e suas frações foram consideradas como contaminante.

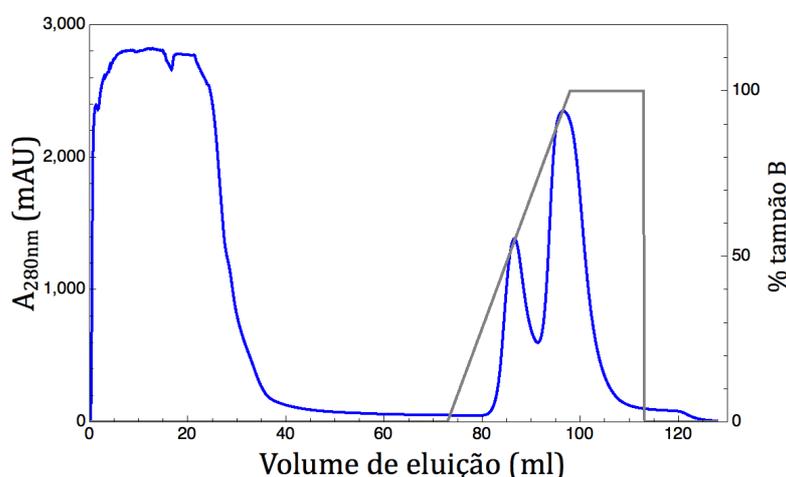


Figura 3.2.1 – Cromatografia de afinidade a níquel (AF) a partir do extrato celular bruto; em azul: cromatograma de absorção de luz UV (280nm); em cinza: porcentagem de tampão B.

Na etapa 2, as frações selecionadas na etapa 1 foram aplicadas na coluna de dessalinização *HiPrep 26/10* (GE Healthcare) previamente ambientada com o tampão *desalting* (Tris-HCl 50mM pH 7,45; NaCl 50mM, MgSO₄ 5mM). Essa etapa visa reduzir a concentração de cloreto de sódio e remover o imidazol do meio que mantém a proteína.

A etapa 3 consistiu da remoção da etiqueta de afinidade e foi executada em alguns dos ensaios de purificação, quando se fez necessário obter a proteína sem a *His-tag*. Nessa etapa, a protease TEV a 20% (p/p) foi utilizada na proteólise da cauda de histidinas. A reação foi feita em tampão Tris-HCl (50mM pH 7,45; NaCl 50mM, MgSO₄ 5mM e DTT 1mM) por 16h a 4°C e nenhum sinal de agregação ou precipitação foi observado. Para separar a SMase D do meio contendo TEV e caudas de histidinas após a clivagem, uma nova cromatografia de afinidade a níquel foi realizada. A protease TEV contém *His-tag* em seu N-terminal e permanece retida, enquanto a proteína clivada elui pela coluna e é recolhida já nas primeiras frações. A protease TEV é produzida no próprio laboratório BioEst do ICB da UFMG a partir do vetor pMHT238Δ. Sua atividade é sítio-específica e reconhece a sequência de resíduos ENLYFQS, ou ENLYFQG, presente entre a etiqueta de afinidade e o início da sequência da proteína recombinante. A clivagem ocorre após o resíduo glutamina [89] e dessa forma, dois resíduos de aminoácidos extras, glicina e histidina, permanecem aderidos ao N-terminal da SMase D. O peso molecular estimado da enzima clivada é de 31.355Da.

Por fim, a etapa 4 consistiu da realização da cromatografia por exclusão molecular (EM) em coluna *HiLoad 16/60 Superdex 200 prep grade* (GE Healthcare) que possui a faixa de separação de 10 a 600kDa. O cromatograma de absorção de luz UV é mostrado na Figura 3.2.2. Para a EM, a coluna foi ambientada em tampão GF (Tris-HCl 50mM pH 7,45; NaCl 50mM, MgSO₄ 5mM, PMSF 1mM) e em seguida as frações selecionadas na etapa anterior, 2 ou 3, foram injetadas. A eluição das proteínas também foi feita com tampão GF. No cromatograma observa-se dois pequenos picos de absorbância para volumes de eluição em torno de 45 e 83mL e um pico de alta absorbância centrado próximo de 97mL. A SMase D é uma proteína de baixo peso molecular e é esperado que ela elua da coluna de EM próximo do fim da lavagem, que acontece em 120mL. A

presença do primeiro pico de absorvância provavelmente se deve a alguma proteína provinda do extrato celular bruto que apresentou afinidade ao níquel na etapa 2, porém é muito maior que SMase D. Já o segundo pico concorda com uma possível fração remanescente do que consideramos contaminante, também na etapa 2. Acreditamos que o terceiro pico corresponde a SMase D, portanto, suas frações foram coletadas e armazenadas a 4°C enquanto as demais foram descartadas.

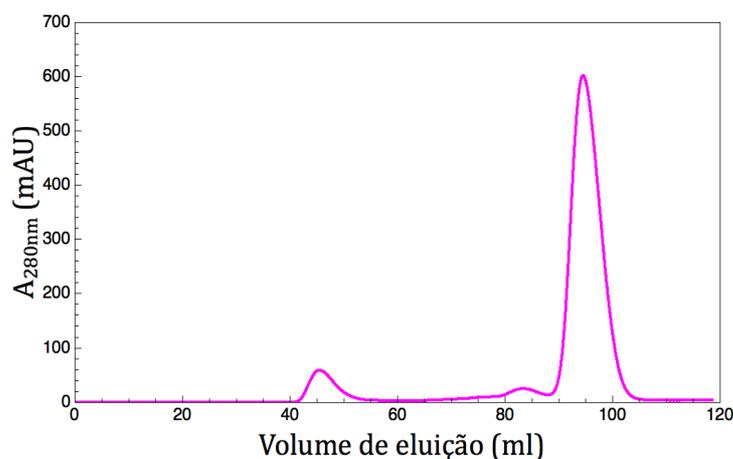


Figura 3.2.2 – Cromatografia por exclusão molecular (EM) das frações selecionadas na AF; cromatograma de absorção de luz UV (280nm) da purificação de 6xHis-SMase D;

Sempre que necessário, as frações de proteína obtidas nas etapas de purificação foram concentradas separadamente por ultrafiltração em *Vivaspin* (GE Healthcare) com limite de exclusão molecular de 10kDa.

Também as alíquotas do eluato das cromatografias de afinidade e de exclusão molecular foram submetidas à eletroforese em gel de poliacrilamida-SDS (Figura 3.2.3). Nele é possível verificar a presença das bandas de aproximadamente 33kDa, destacadas com a seta em vermelho, consistentes com o monômero de 6xHis-SMase D, cujo peso molecular é de 33.282Da. Essas bandas estão presentes nas colunas das alíquotas de 4 horas de expressão (4h), do eluato da cromatografia de afinidade (AF) e do eluato da cromatografia por exclusão molecular (EM) e, como esperado, essa banda não está presente na coluna de zero hora de expressão (0h). Isso confirmou o sucesso na expressão de 6xHis-SMase D e também das etapas de purificação, AF e EM.

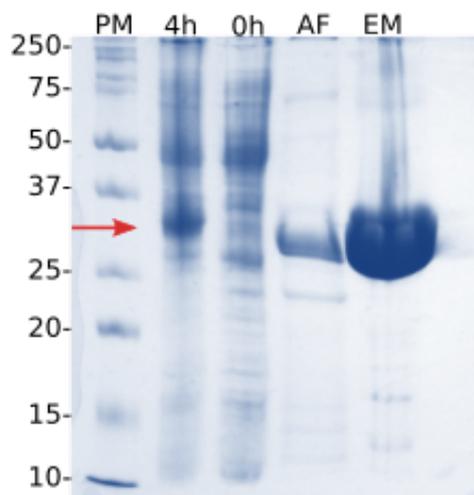


Figura 3.2.3 – Perfil eletroforético de proteínas totais da bactéria *E. coli* BL-21 transformadas com o vetor pET28a-TEV: antes (0h) e 4h após a indução por IPTG (4h), após cromatografias por afinidade (AF) e exclusão molecular (EM). A seta vermelha indica as bandas referentes à 6xHis-SMase. Os valores à esquerda, em kiloDaltons (kDa) são os pesos moleculares de referência (PM). Gel de poliacrilamida-SDS 12% corado com azul de Coomassie.

Quantificação da concentração proteica e rendimento

A dosagem da proteína é feita através de medida de absorvância a 280nm em espectrofotômetro *Nanodrop*[®]. É possível prever o coeficiente de extinção da proteína em estudo a partir da sua estrutura primária através da equação [90]:

$$\varepsilon_{280} = \#Trp \times 5500 + \#Tyr \times 1490 + \#cistina \times 125 \quad 3.2.1$$

onde ε_{280} é o coeficiente de extinção molar para luz UV com $\lambda = 280nm$, $\#Trp$ corresponde ao número de resíduos de triptofano e similarmente $\#Tyr$ ao número de resíduos de tirosina e $\#cistina$ ao número de pontes dissulfeto presentes na proteína.

Para então calcular a concentração molar da proteína, utiliza-se a fórmula de Lambert-Beer:

$$A = \varepsilon \times l \times c \quad 3.2.2$$

onde A é a absorvância, ε é o coeficiente de extinção molar da proteína em $M^{-1}cm^{-1}$, l é o caminho óptico em cm^{-1} e c é a concentração molar da proteína.

Os valores de ε para a proteína nas duas formas, fusionada ou não à His-tag, foram obtidos utilizando-se o programa *ProtParam* [91] hospedado no servidor proteômico *ExPASy* [92]. Para 6xHis-SMase D, $\varepsilon = 65890 M^{-1}cm^{-1}$, e após a clivagem da His-tag, $\varepsilon = 58900 M^{-1}cm^{-1}$.

As dosagens foram feitas após cada etapa de purificação e sempre que necessário. A partir delas vimos que o rendimento da produção e purificação da SMase D foi, em média, de 8mg da enzima para cada litro de meio de cultura.

Espalhamento dinâmico de luz

A caracterização do nível de polidispersividade e do estado de agregação da proteína, bem como estimativas da massa molecular das partículas em solução foram realizadas utilizando dados de espalhamento dinâmico de luz (DLS) realizados em aparelho *Zetasizer Nano ZS-90* (Malvern) a 25°C. Para retirar possíveis impurezas ou partículas insolúveis presentes nas amostras, elas foram submetidas a centrifugação a 13000g por 10 minutos a 4°C. Utilizou-se cubeta de quartzo cuja espessura propiciava um caminho óptico de 10mm e a análise dos dados foi feita utilizando o programa *Malvern Zetasizer v. 6.01*. Os ensaios de DLS foram realizados após a etapa 4 de purificação. Na Figura 3.2.4, vemos as distribuições dos tamanhos das partículas em solução para a SMase D, com e sem His-tag. Nos gráficos, a distribuição de tamanhos é mostrada em relação à intensidade de espalhamento (linha sólida) e em relação à massa (linha tracejada), ambas dadas em porcentagem do total. Por porcentagem de massa, refere-se à contribuição de cada população de partículas para a intensidade de espalhamento da luz.

Para ensaios com a SMase D com His-tag, a curva de intensidade indica dois grupos de partículas, com aproximadamente 5nm e 615nm de diâmetro cada; para ensaios com a SMase D sem His-tag, a mesma curva indica quatro grupos de partículas, com aproximadamente 3, 14, 255 e 4800nm de diâmetro. Apesar da

curva de intensidade indicar a presença de mais de um grupo de partículas, a curva de porcentagem de massa indica que as amostras analisadas eram compostas quase exclusivamente pelas partículas de 5 e 3nm de diâmetro, indicando uma amostra monodispersa com alto grau de pureza. A estimativa da massa molecular a partir do diâmetro, assumindo proteínas globulares, indicou proteínas de peso molecular de aproximadamente 27-30kDa, correspondente ao peso molecular dos monômeros de SMase D.

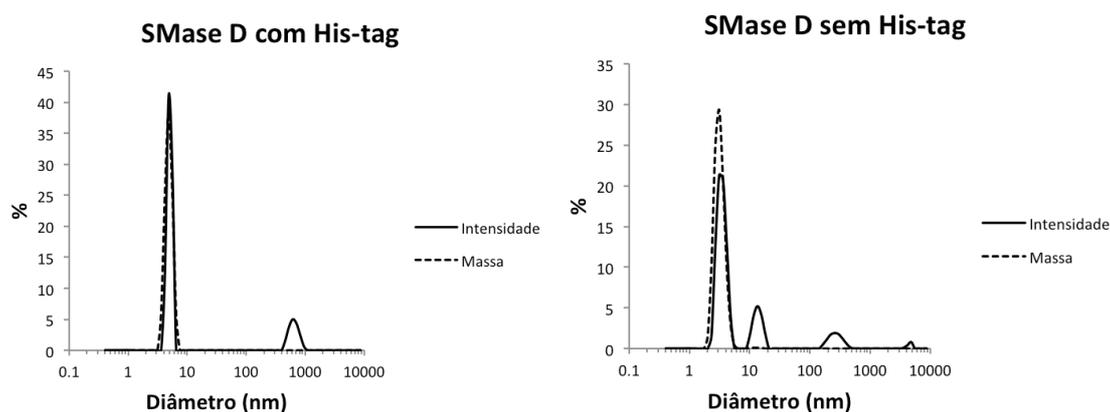


Figura 3.2.4 – Distribuição de tamanhos de partícula para SMase D recombinante obtido por DLS

Dicroísmo circular

Ensaio de espectroscopia de dicroísmo circular (CD) foram realizados utilizando um espectropolarímetro *JASCO J-815* (Jasco Corporation). Espectros de CD no UV distante (190-260nm) foram registrados para amostras a uma concentração de proteína de 0.3mg/mL em tampão, para temperaturas no intervalo de 20 a 90°C com incrementos de 10°C. A velocidade de varredura foi de 100nm/min com passo de 1nm e 5 acumulações por digitalização e a mesma cubeta de quartzo de 1mm de caminho óptico foi utilizada ao longo de todas as varreduras. A correção dos espectros foi realizada subtraindo-se o espectro da solução tampão. Os espectros de CD corrigidos estão mostrados na Figura 3.2.5a.

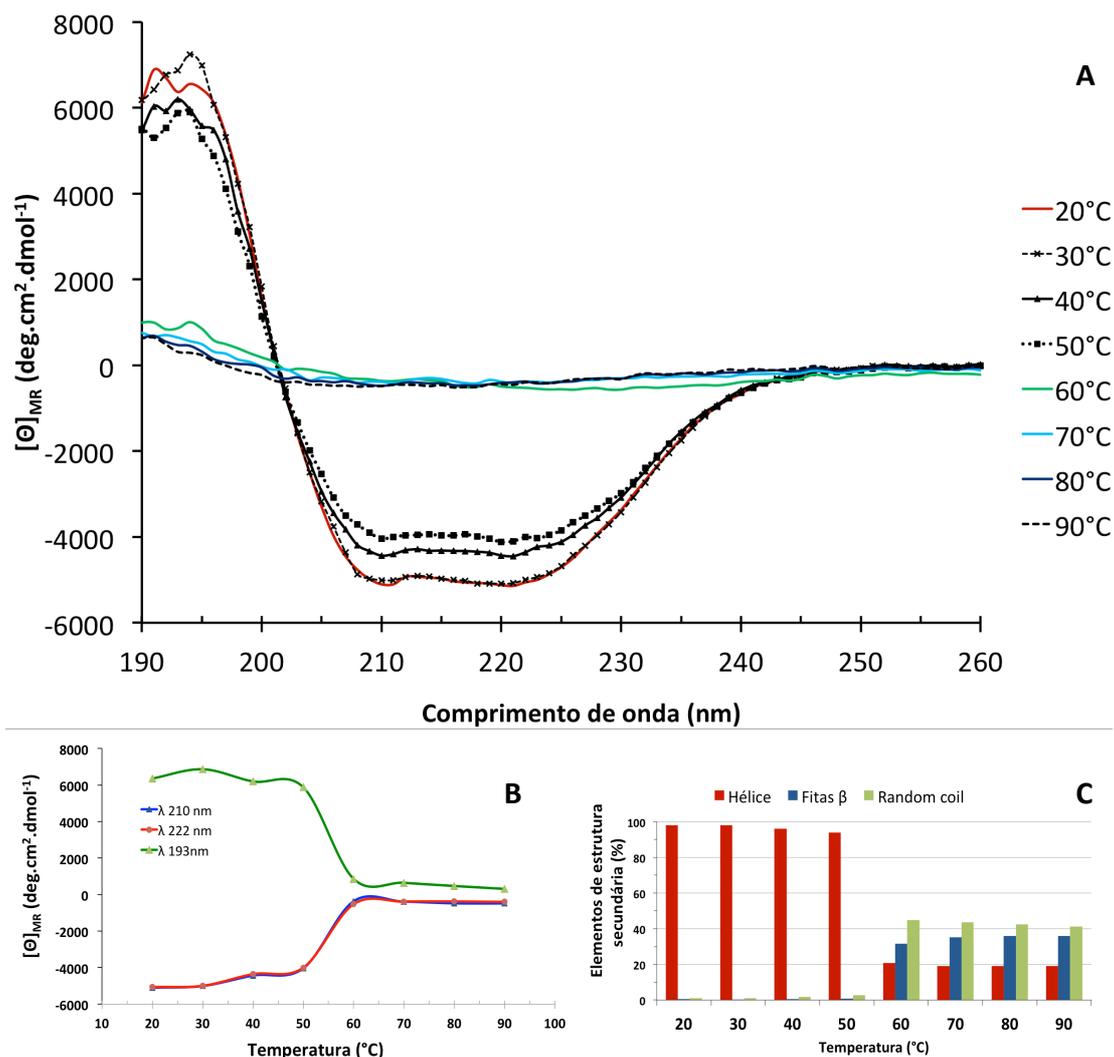


Figura 3.2.5 – Resultados dos ensaios de dicroísmo circular para a SMase D. **A** – Elipticidade molar residual *versus* comprimento de onda (190 a 260nm) para as temperaturas de 20 a 90°C. **B** – Monitoramento da elipticidade molar residual para comprimentos de onda fixos, 193nm (verde), 210nm (azul) e 222nm (vermelho) em função da temperatura. **C** – Conteúdo de elementos de estrutura secundária (%) em função da temperatura.

Em uma análise inicial, observa-se a mudança brusca do padrão da elipticidade molar residual, θ , quando o sistema vai de 50°C para 60°C, indicando uma clara mudança estrutural da proteína. Para as temperaturas entre 20°C e 50°C, observa-se um espectro típico de α -hélices, com contribuições positivas de θ em torno de 193nm e negativas em torno de 208nm e 222nm [35]. Para as temperaturas entre 60°C e 90°C, os espectros apresentam um perfil característico de proteínas desenoveladas, que sugere a drástica perda da estrutura secundária helicoidal.

Essa perda pode ser confirmada ao monitorarmos a elipticidade residual para comprimentos de onda fixo, como mostrado na figura Figura 3.2.5b para 193nm, 208nm e 222nm, onde verificamos a perda dos valores de θ característicos das α -hélices. O resultado da deconvolução desse espectro, onde utilizou-se o programa *CDNN 2.0* [31], é mostrado na Figura 3.2.5c e sugere uma composição essencialmente helicoidal (98% de α -hélices, 1% de fitas- β e 1% de estruturas randômicas, ou *random coil*) para temperaturas abaixo de 60°C, e para as temperaturas mais altas, o conteúdo de α -hélices é reduzido para aproximadamente 20%. Esses resultados mostram que a proteína aparenta ter sua estrutura secundária quase exclusivamente composta α -hélices e que ela mantém sua estabilidade até 50°C, quando o desenovelamento é favorecido. Esse resultado nos chama atenção uma vez que as demais estruturas de SMases D apresentam estruturas majoritariamente compostas por folhas- β .

Cristalização

Dispondo de proteína pura e monodispersa, iniciaram-se os ensaios de cristalização. O passo inicial foi o teste para encontrar a melhor concentração a ser utilizada nas varreduras de condições de cristalização. O kit *PCT™ Pre-Crystallization Test* (Hampton research) foi utilizado e os resultados indicaram que as melhores concentrações da SMase D para ensaios de cristalização giravam em torno de 5mg/mL de proteína, para a qual observou-se um leve precipitado granular.

Uma série de ensaios de cristalização foram realizados para a SMase com *His-tag* (concentrações: 1, 2,5, 5 e 9mg/mL) e sem *His-tag* (concentração: 5mg/mL), sendo que os primeiros ensaios foram feitos para ambas e, uma vez tendo melhores resultados para a enzima com a cauda, os demais ensaios prosseguiram apenas para a SMase D não clivada. Todos os ensaios foram feitos de forma automatizada utilizando as instalações do Laboratório Automatizado de Cristalização de Macromoléculas do Laboratório Nacional de Biociências (ROBOLAB, LNBio) em Campinas, SP. A técnica de cristalização adotada foi a difusão de vapor em gota sentada em placas de 96 poços. As varreduras de

condições de cristalização foram feitas utilizando os kits *Crystal Screen HT* e *Salt Rx*, da Hampton research, *Wizard I* e *II* e *Precipitant Synergy*, da Emerald Biosystems, *Pact Suite* e *JCSG+* da Qiagen, totalizando 544 condições distintas.

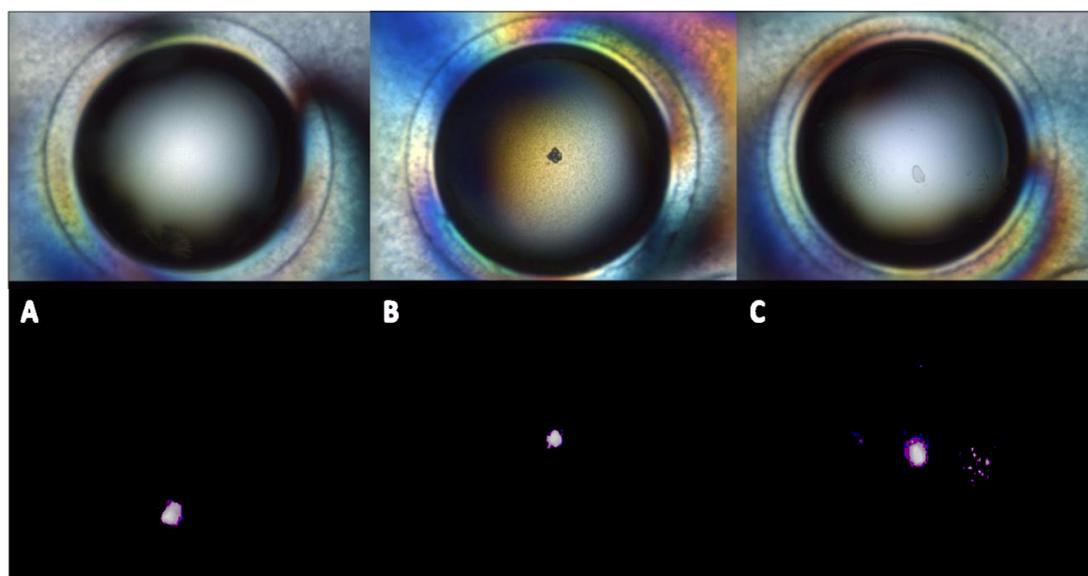


Figura 3.2.6 – Condições que apresentaram formação de cristais. **A** - Nitrato de Potássio 0.2M + PEG 3350 20%p/v. Proteína com His-tag a 2,5 mg/mL. **B** - 2-Metil-2,4-pentanediol 30%v/v + PEG8000 15%p/v + Cloreto de Cálcio 0.1M + Acetato de Sódio 0.1M pH5.5. Proteína com His-tag a 5 mg/mL. **C** - BIS-TRIS propano 0.1M pH7.0 + Citrato trissódico 1.2M. Proteína com His-tag a 5 mg/mL. As imagens inferiores foram obtidas sob exposição à luz UV.

Imagens dos poços de cristalização foram capturadas automaticamente ao longo de dois meses (0, 1, 2, 3, 5, 8, 13, 21, 34 e 55 dias após a montagem das placas) e posteriormente quando requerido. As imagens foram capturadas com luz visível e com luz ultra violeta e eram acessadas remotamente. De todas as condições testadas, três chamaram atenção pela formação de pequenos cristais de proteína: **A** - Nitrato de Potássio 0.2M + PEG 3350 20%p/v. Proteína com His-tag a 2,5 mg/mL; **B** - 2-Metil-2,4-pentanediol 30%v/v + PEG8000 15%p/v + Cloreto de Cálcio 0.1M + Acetato de Sódio 0.1M pH5.5. Proteína com His-tag a 5 mg/mL; **C** - BIS-TRIS propano 0.1M pH7.0 + Citrato trissódico 1.2M. Proteína com His-tag a 5 mg/mL. As imagens dos poços capturadas com luz visível e luz UV podem ser vistas na Figura 3.2.6.

Os cristais da condição A foram submetidos a testes de difração por monocristais na linha MX2 do LNLS, entretanto não foi possível detectar nenhum máximo de difração. Posteriormente as três condições que se apresentaram

promissoras foram então amplificadas manualmente utilizando-se a técnica de difusão de vapor em gota suspensa em placas de 24 poços. As placas foram deixadas nas instalações do Laboratório Nacional de Biociências (LNBio) e serão monitoradas semestralmente pela equipe do laboratório BioEst (ICB/UFMG).

Modelagem

O servidor I-TASSER [93] foi utilizado na previsão da estrutura da SMase D partindo de sua estrutura primária. Ambos essencialmente alinham a estrutura primária da proteína em estudo com as de várias outras proteínas na busca de elementos comuns. Não observamos uma grande diferença entre os resultados, que na média indicaram uma estrutura com 35% hélices, 14% fitas e 51% de estruturas randômicas (*random coil*). O servidor I-TASSER também forneceu a previsão da estrutura tridimensional da SMase D, que está ilustrado na Figura 3.2.7. Observamos o barril TIM¹¹ característico das proteínas da família com o sítio ativo preservado no centro.

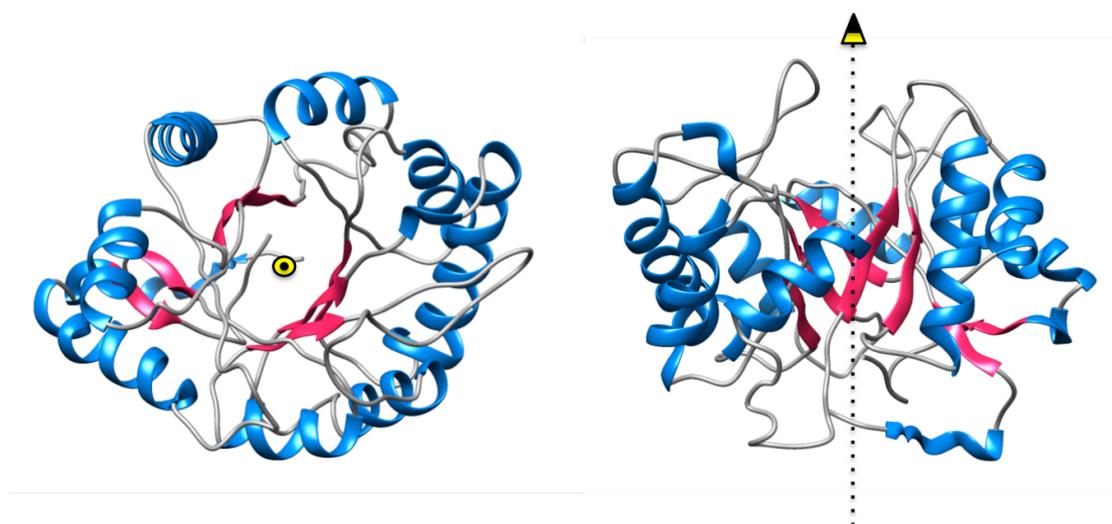


Figura 3.2.7 – Estrutura tridimensional da SMase D prevista pelo servidor I-TASSER vista de duas perspectivas. As folhas- β (rosa) formam a parte interior do barril TIM, e as α -hélices (azul) compõem a superfície. O eixo de referência das figuras é mostrado saindo do plano do papel na figura da esquerda e paralelo ao plano do papel na figura da direita.

¹¹ Enovelamento proteico com oito α -hélices e oito fitas- β

3.3 Conclusão

O estudo de proteínas da família das Esfingomielinases D é de grande importância na compreensão dos mecanismos de patogenicidade dos organismos aos quais elas pertencem. O número reduzido de estruturas de enzimas dessa família depositadas no PDB reforça ainda mais essa importância e motivou a tentativa de determinação estrutural da SMase D de *C. ulcerans*.

Nesse trabalho, a obtenção da proteína recombinante foi bem sucedida, apresentando um rendimento médio razoável de 8mg para cada litro de meio de cultura. Os cromatogramas das purificações mostraram a presença de duas frações de proteína, uma instável, que chamamos de contaminante, e uma fração de proteína estável em quantidade e adequada para os ensaios de cristalização quando analisada por DLS, onde ela se mostrou monodispersa. Os ensaios de DLS também mostraram que a proteína produzida através dos protocolos utilizados se apresentou pura e que o seu estado de agregação estável é monomérico. Os experimentos de espectroscopia de CD nos permitiram concluir que a SMase D é estruturalmente estável para temperaturas abaixo de 30°C, mantém sua estrutura secundária até 50°C e, a partir daí, o aumento da temperatura leva a uma perda drástica do conteúdo de estrutura secundária, que resulta na desnaturação da enzima. A deconvolução do espectro, entretanto, indicou que a SMase D é composta majoritariamente por α -hélices, o que discorda da previsão de estrutura secundária a partir da estrutura primária pelo servidor *I-TASSER*. As demais enzimas da família da SMase D apresentam o barril-TIM com o sítio ativo conservado no centro, como previsto. Isso chama atenção para o resultado obtido. Apesar do alinhamento de sequências da SMase D de *C. ulcerans* com as demais PLD's da família mostrar uma baixa similaridade entre suas estruturas primárias, esperaríamos que a forma das proteínas da mesma família fossem similares. Essa dúvida somente será esclarecida com a determinação da estrutura tridimensional da proteína.

Dentre todas as condições de cristalização avaliadas nos ensaios, apenas três apresentaram a formação de pequenos cristais. Os cristais formados não se mostraram adequados para experimentos de difração de raios X e as condições

foram recentemente amplificadas. Ainda não foi possível observar a formação de novos cristais, e isso reforça a necessidade de aumentar o número de tentativas, varrendo ainda mais parâmetros termodinâmicos e/ou testando outras técnicas de cristalização.

Em suma, os objetivos do trabalho com a SMase D de *C. ulcerans* foram parcialmente atingidos, uma vez que condições de cristalização foram encontradas. Entretanto ainda é necessário que elas sejam otimizadas levando à formação de cristais adequados para experimentos de difração de raios X e à determinação da estrutura da enzima.

Complexo NahK/NahL

4.1 Introdução

Os compostos aromáticos monocíclicos, como benzeno, tolueno e xileno, e os hidrocarbonetos aromáticos policíclicos (HAPs) como naftaleno, antraceno e pireno, estão entre os contaminantes mais comuns e tóxicos dentre os hidrocarbonetos totais do petróleo [94]–[97]. Isso faz com que a indústria e a comunidade científica se atentem para medidas a serem tomadas para a redução da concentração desses compostos em áreas contaminadas.

Os processos que auxiliam na diminuição da concentração dos poluentes do ambiente são denominados processos de remediação e podem ser físicos, químicos ou biológicos. Dentre os processos químicos e físicos estão a remoção dos poluentes via sistemas de contenção e coleta, via processos de adsorção por outros materiais, como carvão ativado, e via processos de aquecimento. A biorremediação, por sua vez, engloba diversas técnicas que fazem uso de processos de degradação biológicos para transformar as moléculas de contaminantes em moléculas atóxicas ou menos tóxicas. A biorremediação é uma tecnologia limpa e já vêm sendo utilizada em escala comercial no tratamento de resíduos e na remediação de áreas contaminadas [98]– [100].

A via de degradação do naftaleno

Diversas espécies de fungos e bactérias vem sendo estudadas desde a década de 1950 devido a sua capacidade de consumir diferentes tipos de HAPs e novas vias de degradação vem sendo elucidadas [101]–[104]. Particularmente, as bactérias *Pseudomonas putida*, comumente encontradas em solos contaminados, foram

isoladas e estudadas devido a sua capacidade de degradar os HAPs de forma eficiente [101]. Elas ganham atenção devido à presença de um plasmídeo de 83 mil pares de base, NAH7 [105], que codifica 20 proteínas que juntas são capazes de degradar completamente o naftaleno. O naftaleno é um HAP simples formado por dois anéis aromáticos fundidos que é altamente solúvel em água, sendo um dos primeiros dentre os HAPs a atingir lençóis freáticos em caso de contaminação do solo.

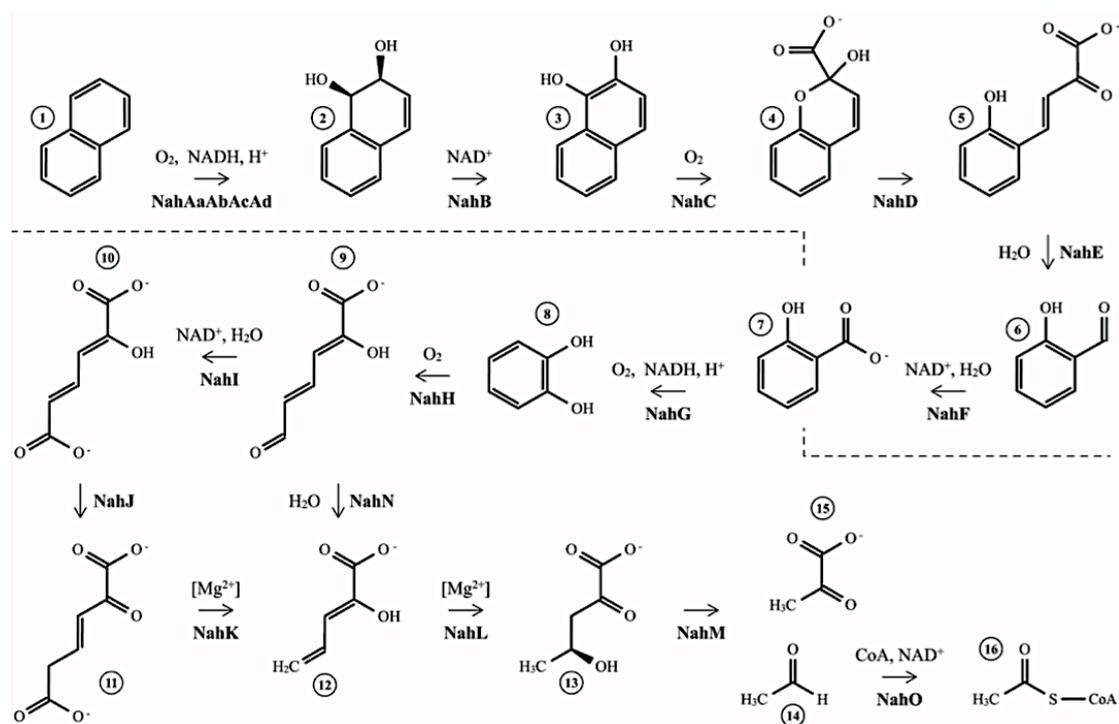


Figura 4.1.1 – Via de degradação do Naftaleno pelas enzimas de *Pseudomonas putida*. As vias superior e inferior estão divididas pela linha pontilhada. Os compostos numerados ao longo da via são: (1) naftaleno; (2) cis-1,2-dihidronaftaleno-1,2-diol; (3) 1,2-dihidroinaftaleno; (4) 2-hidroxi-2H-cromeno-2-carboxilato; (5) trans-o-hidroxibenzilidenopiruvato; (6) salicilaldeído; (7) salicilato; (8) catecol; (9) 2-hidroxi-2,4-pentadienoato semialdeído; (10) 2-hidroxi-2,4-pentadienoato; (11) 2-oxo-3-hexenedioato (4-oxalocrotonato); (12) 2-hidroxi-2,4-pentadienoato; (13) 2-oxo-4-hidroxipentanoato; (14) piruvato; (15) acetaldeído; (16) acetil-CoA.

A via de degradação do naftaleno está ilustrada na Figura 4.1.1. Ela é dividida em duas vias, superior e inferior, nas quais acontece a degradação do primeiro e do segundo anel aromático do naftaleno, respectivamente. Na via superior atuam o complexo proteico NahAaAbAcAd e as enzimas NahB, NahC, NahD, NahE e NahF. Na via inferior atuam as enzimas NahG, NahH, NahI, NahJ, NahK, NahL, NahM e NahO.

O complexo NahK/NahL

Como pode ser visto na Figura 4.1.1, a quinta e a sexta reação na via inferior de degradação do naftaleno são realizadas pelas enzimas NahK e NahL, respectivamente. NahK é uma 4-oxalocronato descarboxilase com 28.5k Da de massa molecular cuja estrutura primária é composta pelos 264 resíduos de aminoácidos da sequência:

**MNRTLREQVLALAEHIENAELNVHDIGKVTNDFPEMTFADAYDVQWEIRRRKEARGNKIVG
LKMGLTSWAKMAQMGVETPIYGFLADYFSVPDGGVDCSKLIHPKIEAEISVVKAPLHGP
CHLGDVIAAIDYVIPTVEVIDSRYENFKFDPIVVADNASSTRFITGGRMASLEEVDLRTLGV
MEKNGEVVELGAGAAVLGHPLSSVAMLANLLAERGEHIPAGTFIMTGGITAAPVAPGDNIT
VRYQGLGSVSARFI**

Por sua vez, NahL, denominada vinilpiruvato hidratase, tem 28kDa de peso molecular e sua estrutura primária é composta pelos 261 aminoácidos da sequência:

**MDQILINELGDELYQAMLNREAIAPLTERGFAISIDDAYHISLRMLERRLAAGEKIIGKIGVTS
KAVQNMLGVYQPDFGYLTDAMVFNSGEAVPISQRLMQPKAEGEIAFILKKDLIGPGVTNADV
LAATECVMPCFEIVDSRIRDWKIKIEDTVADNASGLFVLGDQAVSPRQVDLVTGMLVEKN
GQLLSTGAGAAALGSPVNCVAWLANTLGRFGIPLKAGEVILSGSLVPLEPVKAGDFMRVEIG
GIGSASVRFT**

Essas duas enzimas foram foco do estudo da tese de doutoramento do Dr. Samuel Leite Guimarães no Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas (ICB) da Universidade Federal de Minas Gerais (UFMG) em 2015 [106]. Em seu trabalho, o Dr. Guimarães descreveu a expressão, a purificação e a cristalização de NahK bem como de seu complexo com NahL (NahK/NahL). Experimentos de difração de raios X por monocristais (XRD) permitiram a determinação da estrutura tridimensional da enzima NahK e, junto a experimentos de espalhamento de raios X de baixo ângulo (SAX), também a elucidação da estrutura do complexo NahK/NahL e a determinação da estrutura tridimensional da enzima NahL. Como registrado na tese, a oligomerização das enzimas forma um heterodecâmero (Figura 4.1.2) no qual elas estão dispostas de

forma a exibir suas cavidades (sítios ativos) uma pra outra, permitindo que o produto instável de NahK (molécula 12 da Figura 4.1.1) rapidamente alcance o sítio ativo de NahL. Os dados de difração de raios X obtidos por Dr. Guimarães, após passarem pelo processo de faseamento por substituição molecular, ainda apresentavam regiões altamente desordenadas, ou seja, regiões onde a densidade eletrônica não era bem definida, o que não permitiu a determinação de um monômero do heterododecâmero, conforme indicado na Figura 4.1.2.

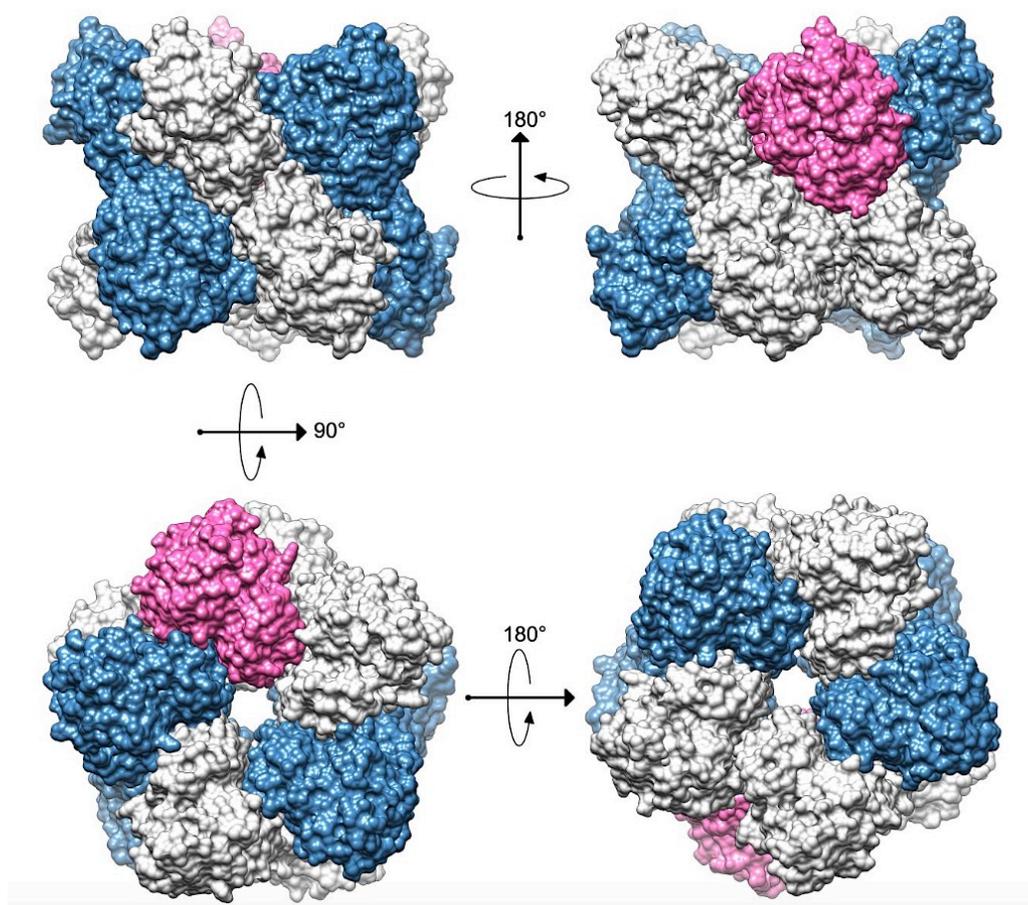


Figura 4.1.2 – Modelo de superfície das subunidades formadoras dos heterododecâmeros de NahK (azul) e NahL (cinza). O monômero rosa não pode ser determinado. Fonte: [106]

Nas próximas seções discutiremos o trabalho feito com os dados de difração de raios X e a tentativa de resolver a desordem e elucidar a estrutura tridimensional completa dos decâmeros que formaram o cristal previamente investigado.

4.2 Procedimentos e discussão

Reprocessamento de dados de XRD

Um único conjunto de dados de difração de raios X de um cristal do complexo NahK/NahL foi coletado na linha MX2 do Laboratório Nacional de Luz Sincrotron (LNLS), como detalhado na tese de doutoramento de Dr. Guimarães. O grupo de espaço que descreve a simetria do cristal foi determinado como sendo $P2_12_12_1$ e a célula unitária determinada apresentava os seguintes parâmetros de rede: $a = 100,41 \text{ \AA}$, $b = 219,96 \text{ \AA}$, $c = 310,36 \text{ \AA}$; e $\alpha = \beta = \gamma = 90^\circ$. O processamento inicial das 1100 imagens de difração obtidas foi feito utilizando o programa *HKL2000* [107] para os limites de resolução $47,66\text{-}2,49\text{\AA}$, o que permitiu a indexação de 985.382 reflexões. Um resumo das estatísticas da medida é mostrado na Tabela 5.2.1.

Na tentativa de extrair mais informação a partir do único conjunto de dados disponível, o reprocessamento das imagens foi feito utilizando o software XDS [59], [60]. De fato foi possível estender a resolução dos dados até 2.20\AA , como pode ser visto na Figura 4.2.1, que mostra a ampliação de uma região próxima ao novo limite de resolução. A presença de alguns picos de interferência construtiva é evidente, apesar do novo limite se estender além do limite horizontal do detector, que coincide exatamente com o círculo de resolução de 2.49\AA .

As etapas de reprocessamento foram realizadas sempre em duplicata, impondo os limites de resolução estendida e, por vias de comparação, os limites utilizados anteriormente por Dr. Guimarães. A comparação entre as estatísticas dos três processamentos de dados é mostrada na Tabela 4.2.1. Para a resolução estendida, o número de reflexões indexadas aumentou para 1.280.561 (295.179 reflexões a mais, 29,9%), que corresponde a 98.102 (41,8%) reflexões únicas (reflexões não equivalentes por simetria) a mais. Para a mesma resolução o aumento foi de 50.038 (5,1%) e 1605 (0,7%) respectivamente. O aumento significativo do número de reflexões únicas ao estendermos a resolução é coerente com o aumento da área considerada na busca de reflexões. Todavia, ao considerarmos os mesmos limites de resolução utilizados anteriormente, vemos

que o número de reflexões indexadas também muda, uma vez que os softwares *XDS* e *HKL2000* utilizam metodologias diferentes na avaliação das imagens de difração, mas não se observa um aumento tão significativo de reflexões únicas.

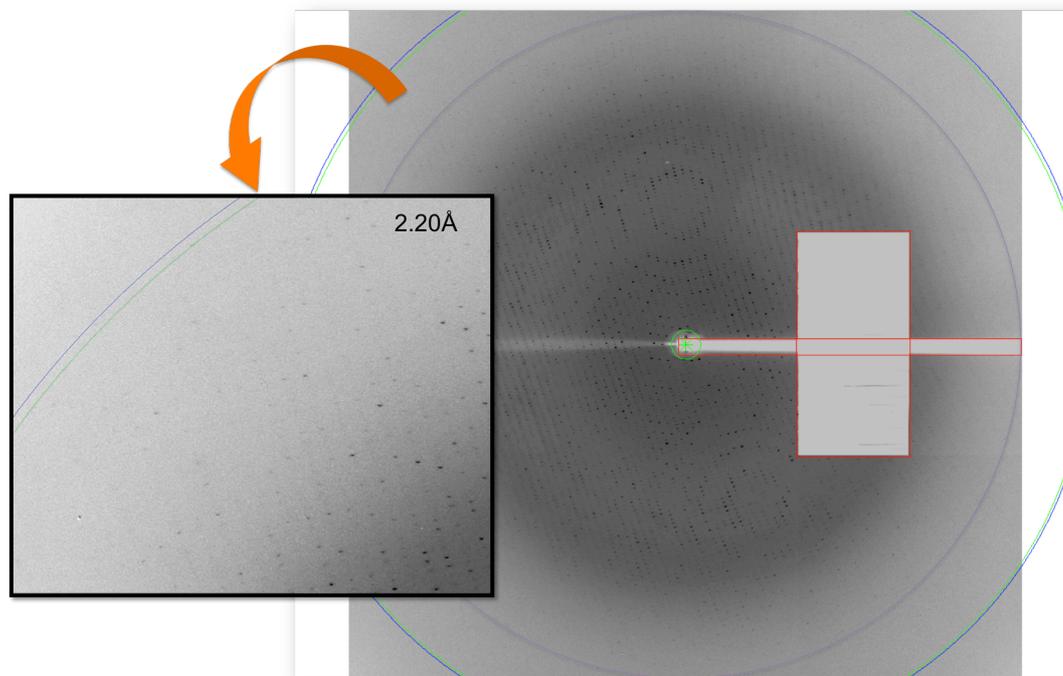


Figura 4.2.1 – Ampliação de uma região próxima à circunferência correspondente à resolução de 2.20Å (círculo verde). A circunferência correspondente à resolução de 2.49 Å é mostrado em roxo. A região delimitada em vermelho corresponde à sombra do *beamstop* e a uma região onde o detector apresentou problemas na detecção de radiação.

Os demais parâmetros da estatística do processamento não apresentaram mudanças significativas, mas é possível verificar a diminuição de todos os demais parâmetros de qualidade. Certamente, medir apenas uma fração das últimas camadas de resolução contribui negativamente para as estatísticas, entretanto foi necessário estabelecer um compromisso entre aumentar a quantidade de reflexões consideradas e diminuir a qualidade total do processamento. Dado que não houve nenhuma mudança que tornasse o conjunto de dados inadequado para prosseguir com a solução da estrutura, todas as demais etapas foram realizadas com os dados processados com resolução estendida e, para fins de comparação, também com os dados reprocessados para a mesma resolução.

Tabela 4.2.1 – Estatísticas dos processamentos de dados utilizando o softwares *HKL2000* e *XDS*

	HKL2000	XDS	
	Resolução (Å)	47,66-2,49 (2,53-2,49)	47,66-2,49 (2,64-2,49)
Reflexões	985.382	1.035.420	1.280.561
Reflexões únicas	234.835	236.440	332.937
$R_{\text{merge/meas}}(\%)*$	7,4 (30,0)	6,5 (34,4)	7,9(94,8)
$\langle I/\sigma(I) \rangle$	13,77 (3,42)	15,40 (3,97)	11,57 (1,22)
Completeza (%)	97,8 (81,5)	99,0 (99,2)	96,5 (83,4)

Os valores entre parênteses se referem aos dados da última camada de resolução.

*Para os dados processados com o software *HKL2000*, temos R_{merge} , e para dados processados com o software *XDS*, temos R_{meas} .

Estrutura de NahK/NahL e desordem

Assim como reportado na tese de Dr. Guimarães, o faseamento inicial dos dados de difração de raios X obtidos foi feito através de uma substituição molecular utilizando a estrutura de NahK como modelo e o conjunto de dados processados inicialmente com o software *HKL2000* até resolução igual a 2,49Å (*HKL2000/2,49Å*). A forma do envelope do complexo havia sido elucidada através de experimentos de SAX e auxiliou na proposta de um modelo inicial do heterodecâmero. Os dados de XRD indicaram a presença de 20 monômeros na unidade assimétrica dispostos como dois decâmeros heterogêneos conectados. Entretanto a presença de algumas regiões de eletrônica pobre não possibilitou a determinação da natureza NahK ou NahL de dois desses monômeros (um em cada heterodecâmero) de forma assertiva, mesmo depois de alguns refinamentos. A dificuldade na determinação desses monômeros é melhor

compreendida quando faz-se a superposição das estruturas de NahK e NahL determinadas por Dr. Guimarães. Na Figura 4.2.2 vemos a quase total superposição das respectivas estruturas secundárias, sendo o desvio quadrático médio (RMSD) entre a posição de todos os átomos da cadeia principal das enzimas de apenas 1.32Å. Logo, a principal diferença entre as duas enzimas está nas cadeias laterais dos resíduos alinhados estruturalmente e, nas regiões onde a densidade eletrônica se mostra pobre, não é possível ver detalhes dessas cadeias além da provável posição das cadeias principais.

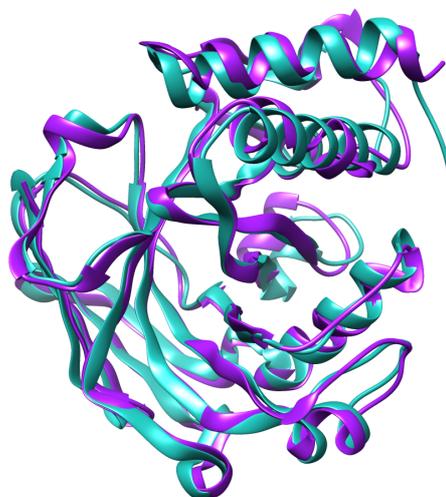


Figura 4.2.2 – Alinhamento estrutural de NahK (ciano) e NahL (roxo).

Esse problema foi então trazido para o Laboratório de Cristalografia do departamento de física da UFMG, onde foram feitas tentativas de retrabalhar os dados de XRD na busca da determinação da natureza dos monômeros não identificados. Primeiramente, lembramos que o mapa de densidade eletrônica é sempre calculado utilizando-se os dados de XRD medidos (intensidades) e as fases calculadas para um modelo estrutural (ver seção 2.4, Difração por cristais 3D). Conseqüentemente, o modelo utilizado tem influência direta nas fases calculadas para o mapa experimental. Para contornar esse problema, calculamos inicialmente um mapa de densidade eletrônica experimental não tendencioso. Para tal, o modelo de busca para uma nova substituição molecular foi manipulado de tal forma que, ao calcular as fases, não houvesse tendenciosidade gerada pela presença de cadeias laterais de NahK ou NahL em nenhum dos monômeros. Posteriormente, novos mapas de densidade eletrônica e de

diferença de Fourier foram calculados para então identificarmos as regiões de densidade eletrônica ao longo das cadeias principais de cada monômero que permitissem a determinação de sua natureza.

A manipulação do modelo de busca consistiu na modificação de todos os monômeros do modelo proposto por Dr. Guimarães de tal forma que foram mantidas apenas as cadeias laterais dos resíduos de aminoácidos idênticos (resíduos KL) que se superpõem perfeitamente quando NahK e NahL são alinhadas estruturalmente, ou seja, quando é feito o alinhamento das sequências baseado na estrutura das enzimas. Todos os demais resíduos foram substituídos por alaninas (ALA), aminoácido cuja cadeia lateral é apenas um carbono. O alinhamento estrutural, ou seja, o alinhamento das sequências baseado na estrutura das enzimas, foi realizado utilizando o programa *GESAMT* [64] da plataforma CCP4 [62]. O novo modelo construído dessa forma será chamado de “KL/ALA” daqui pra frente.

Fez-se uma nova substituição molecular com o software *PHASER* [63] utilizando o modelo KL/ALA e os dados de XRD processados com limites de resolução estendidos para $50,0\text{\AA}-2,20\text{\AA}$ (XDS/2,20 \AA). Um novo mapa de densidade eletrônica foi calculado e superposto ao modelo para análise. Todos os monômeros foram visualmente inspecionados, resíduo por resíduo, através da interface gráfica do software *COOT* [66], [67], na busca de fragmentos de densidade eletrônica que indicassem a presença de cadeias laterais de aminoácidos característicos de NahK e/ou NahL que deveriam ser adicionados ao modelo para que esse explicasse melhor o mapa de densidade eletrônica experimental. Era evidente a presença de várias das cadeias laterais em quase todos os monômeros de KL/ALA. À medida que cada monômero era identificado como NahK ou como NahL, o modelo KL/ALA era editado adicionando-se as cadeias laterais correspondentes. Esses novos modelos aperfeiçoados eram posteriormente refinados utilizando-se o software *REFMAC5* [65] e o conjunto de dados XDS/2,20 \AA . A nomenclatura inicial dada às cadeias de cada monômero foi mantida e está esquematizada na Figura 4.2.3 para referência futura.

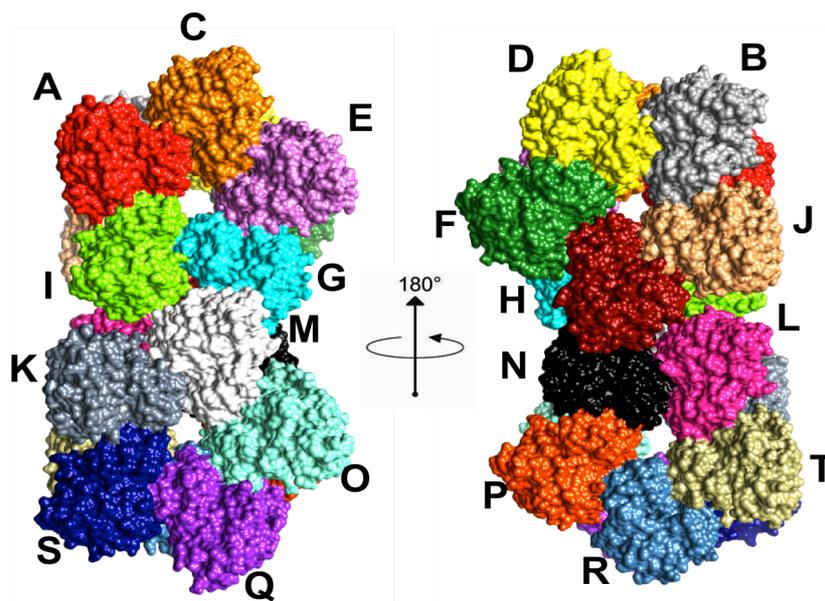


Figura 4.2.3 – Os 20 monômeros que constituem o complexo e sua respectiva nomenclatura dada. Na figura temos os dois decâmeros vistos de cima (esquerdo) e por baixo (direita).

Inicialmente, 19 monômeros foram assinalados NahK ou NahL (todos, exceto G). A evolução dos indicadores de qualidade à medida que se determinava a natureza de cada monômero está mostrado no gráfico da Figura 4.2.4. Nele observamos uma evolução quase que linear de R_{free} e R_{factor} , exceto quando os 16º e 17º monômeros (Q e R, respectivamente) foram assinalados NahL, quando observamos uma quase estagnação dos R 's. Entretanto, a densidade eletrônica nas regiões correspondentes era pobre e o modelo foi novamente modificado mantendo os monômeros Q e R como cadeias de alaninas e resíduos KL.

Os 3 monômeros indeterminados, G, Q e R, demandaram uma análise rigorosa do mapa de densidade eletrônica ao longo de suas cadeias. Para tal, um modelo construído com os 17 monômeros determinados e os outros três ainda como cadeias de alaninas e resíduos KL foi utilizado numa nova substituição molecular com o software *PHASER* [63]. Com o novo mapa gerado, a busca pelos aminoácidos distintos foi retomada. Para auxiliar a busca, utilizamos novamente o programa *GESAMT* [64], que também assinala aos aminoácidos alinhados um nível de dissimilaridade de 0 a 5.

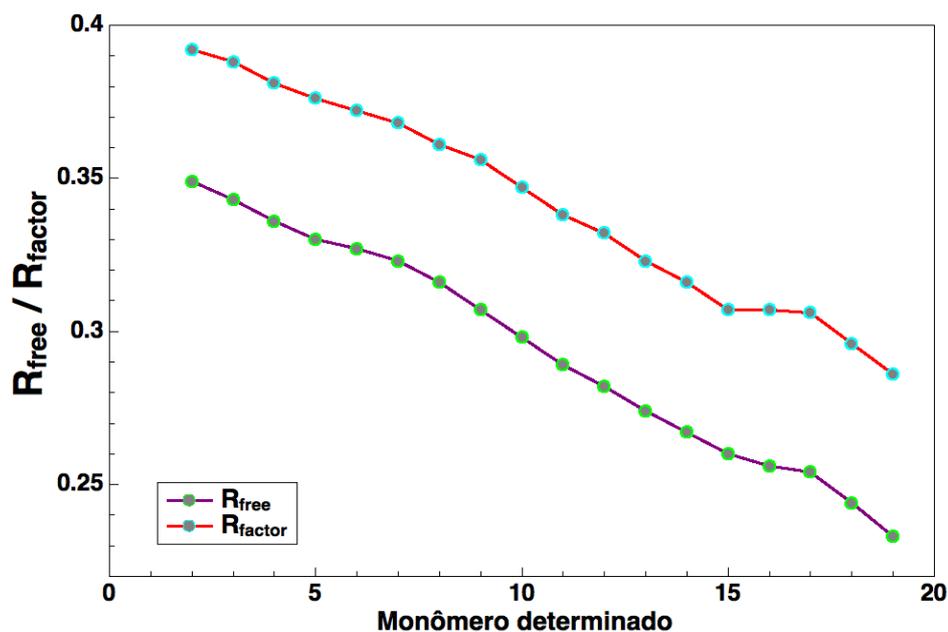


Figura 4.2.4 – Evolução dos indicadores de qualidade do modelo utilizado quando refinado contra os dados experimentais à medida que a natureza dos monômeros foi sendo determinada NahK ou NahL.

Na Figura 4.2.5, vemos o gráfico do número de pares de aminoácidos em cada grupo de similaridade, onde 5 indica aminoácidos totalmente dissimilares e 0, aminoácidos idênticos. Vemos que a maioria dos pares alinhados são de aminoácidos idênticos, mas existe um número considerável de aminoácidos dissimilares. Esses aminoácidos foram meticulosamente analisados na busca por regiões sugestivas no mapa de densidade eletrônica.

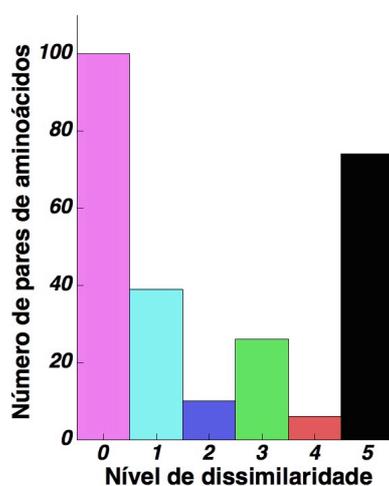


Figura 4.2.5 – Nível de dissimilaridade entre os pares de aminoácidos de NahK e NahL alinhados estruturalmente. Aqui, 0 indica aminoácido idênticos e 5, totalmente dissimilares.

Novamente as cadeias foram inspecionadas visualmente, resíduo a resíduo. Todavia, ao final das comparações dos resíduos, apenas pudemos afirmar que a cadeia nomeada Q possui uma probabilidade maior de ser a enzima NahL. As duas cadeias remanescentes (G e R) permaneceram indeterminadas. A tabela do Apêndice B mostra detalhes dessa análise.

Monômeros G e R e dupla conformação

Desde o início, o contraste ao longo do mapa de densidade eletrônica nos fez questionar o que poderia ter levado à presença de regiões onde é clara a presença dos elementos constituintes da proteína, enquanto em algumas regiões específicas o mapa é dúbio e não permite a construção de um modelo satisfatório para explicar os dados experimentais. Sabemos que uma densidade eletrônica calculada se mostrará pobre quando os dados forem de baixa resolução, quando existir um movimento térmico excessivo dos constituintes do cristal e também por possíveis defeitos ao longo do cristal que prejudiquem a periodicidade. Certamente os dados de XRD aqui utilizados não possuem resolução ótima e podemos atribuir parte dessa indeterminação a isso, apesar de podermos considerá-los de boa qualidade para trabalhos com cristais de proteína. Todavia a desordem observada também poderia advir da ocupação parcial dos sítios das cadeias G e R por ambas enzimas, NahK e NahL, ao longo do cristal, formando heterodecâmeros distintos.

A construção de um modelo que levasse em consideração a possível dupla ocupação dos sítios G e R ao longo do cristal utilizou os 18 monômeros determinados, 10 NahK e 8 NahL, e os outros dois sítios foram ocupados por cadeias de NahK e de NahL superpostas, cada uma com parâmetro de ocupação igual a 0,5. A forma do envelope desse modelo para os dois heterodecâmeros dispostos na unidade assimétrica pode ser vista na Figura 4.2.6.

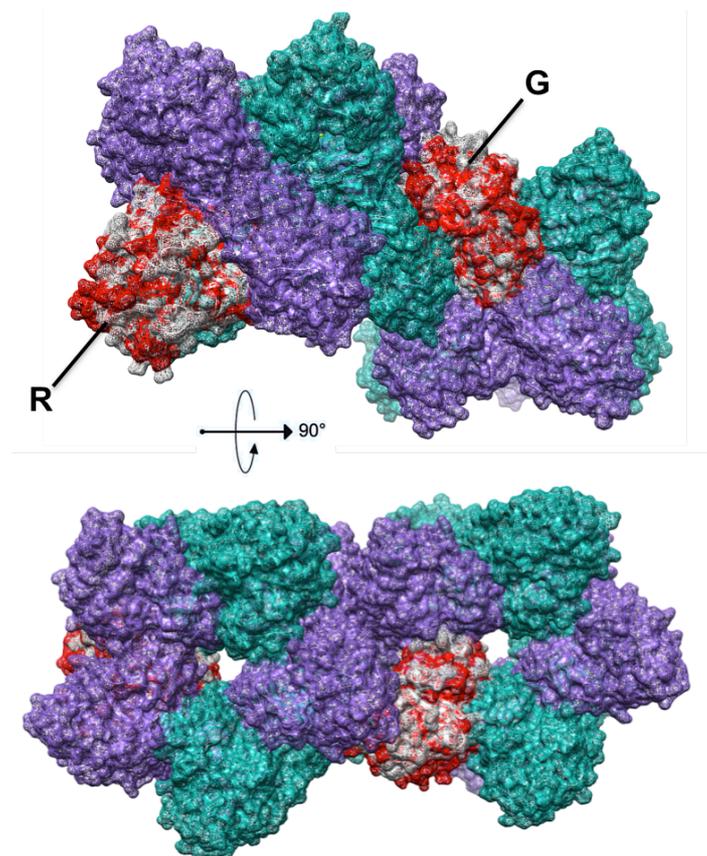


Figura 4.2.6 – Disposição dos dois heterodecâmeros na unidade assimétrica. Em roxo, temos monômeros de NahL e em ciano, monômeros de NahK. Os dois monômeros em vermelho e cinza foram ocupados por cadeias de NahK e de NahL superpostas, cada uma com ocupação igual a 0,5.

Com esse novo e último modelo, novas fases foram calculadas, ou seja, uma nova substituição molecular foi feita, novamente utilizando-se o software *PHASER* [63]. Em seguida o modelo foi refinado em vários ciclos realizados usando os dados XDS/2,20Å à medida que moléculas de água, íons e ligantes que participam da estrutura foram adicionados. Nessa rodada de refinamentos utilizamos novamente o programa *REFMAC5* [65], porém, dessa vez as ocupações de cada enzima, NahK ou NahL, nos sítios G e R entraram como um parâmetro a ser refinado. No entanto, mesmo após 800 ciclos de refinamento, essa nova variável não apresentou mudança. De fato, conforme sugerido pelos desenvolvedores do *REFMAC5*, esse software não é projetado para refinar ocupações de tantos átomos.

A avaliação final dos valores da ocupação das enzimas NahK e NahL nos sítios G e R foi realizada utilizando-se o programa *SHELXL* [108]. 400 ciclos de refinamento foram executados usando os dados XDS/2,20Å. O gráfico na Figura

4.2.7 mostra a evolução do valor da ocupação de NahK ao longo dos ciclos de refinamento que convergiu para 0.45 para o sítio G, e 0.47 para o sítio R. Uma vez que os sítios devem ser ocupados por NahK ou NahL, a soma de suas respectivas ocupações deve ser igual a 1, e dessa forma as ocupações de NahL são 0.55 no sítio G e 0.53 no sítio R.

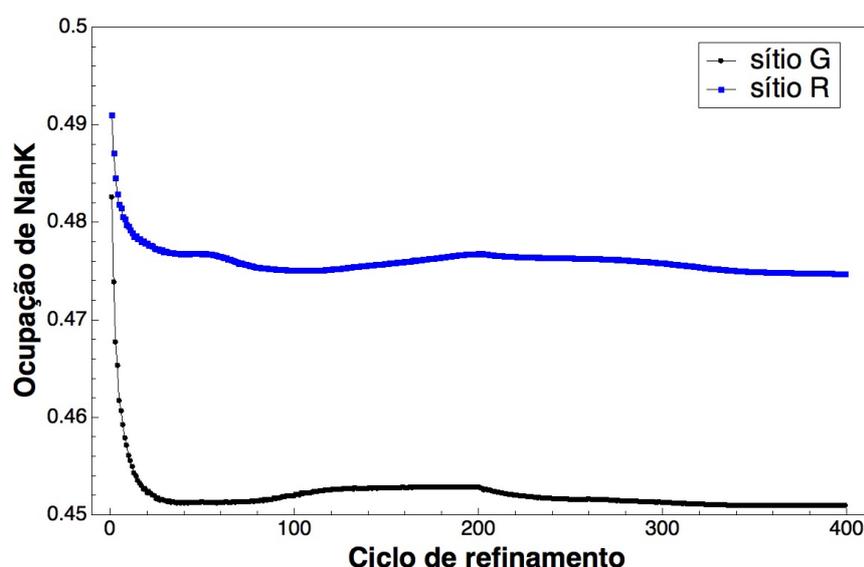


Figura 4.2.7 – Evolução do valor da ocupação de NahK nos sítios G e R ao longo dos ciclos de refinamento.

Utilizando esses valores para a ocupação de NahK e NahL nos sítios G e R, 100 outros ciclos de refinamento foram realizados, dessa vez usando tanto os dados XDS/2,20Å quanto os dados XDS/2,49Å. As estatísticas finais dos refinamentos estão mostradas na Tabela 4.2.2 junto às estatísticas finais do refinamento inicial realizado pelo Dr. Guimarães. Vemos que a mudança nos índices de qualidade do refinamento não é muito grande. De fato, observamos que para o conjunto de dados XDS/2,20Å, os índices de qualidade apontam um ajuste inferior do modelo aos dados experimentais. Entretanto, esse resultado é esperado uma vez que incluímos reflexões menos intensas e camadas de resolução incompletas na tentativa de enxergar mais detalhes no mapa de densidade eletrônica experimental. Para os dados na mesma resolução, os índices são quase idênticos, sugerindo que a desordem observada seja uma característica intrínseca da estrutura.

Tabela 4.2.2 – Estatística final do refinamento do modelo para o complexo NahK./NahL usando os conjuntos de dados XDS/2,20Å e XDS/2,49Å em contraste com a estatística final do refinamento feito por Dr. Guimarães.

Conjunto de dados	HKL2000/2,49Å	XDS/2,49Å	XDS/2,20Å
R _{fator}	0,195	0,193	0,203
R _{free}	0,230	0,255	0,258
RMS ângulos	1,894	1,748	1,763
RMS ligações	0,013	0,014	0,014

4.3 Conclusão

A compreensão de vias de degradação de moléculas por organismos é crucial em sua utilização em processos de biorremediação. As bactérias *Pseudomonas putida* são capazes de degradar o naftaleno, que é um dos hidrocarbonetos totais do petróleo. O naftaleno é uma molécula tóxica e altamente solúvel capaz de atingir os lençóis freáticos em caso de contaminação do solo.

Dentre as 20 proteínas envolvidas na degradação do naftaleno por *P. Putida*, as enzimas NahK e NahL atuam, respectivamente, na quinta e na sexta reação da via inferior (Figura 4.1.1). Elas foram foco do estudo da tese de doutoramento do Dr. Samuel Leite Guimarães no Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas (ICB) da Universidade Federal de Minas Gerais (UFMG) em 2015 [106]. O Dr. Guimarães reportou a estrutura tridimensional de ambas as enzimas e a oligomerização delas em forma de um heterodecâmero, o complexo NahK/NahL, cuja estrutura não pôde ser completamente elucidada a partir de dados de difração de raios X. Foi observado que o mapa de densidade eletrônica experimental construído continha regiões de baixa resolução, o que não permitiu a determinação assertiva na natureza de um dos monômeros de cada decâmero.

Nesse trabalho, reprocessamos os dados de difração de raios X obtidos anteriormente para o cristal do complexo NahK/NahL. Os limites de resolução dos dados puderam ser estendidos de 47,66Å – 2,49Å para 50,0Å – 2,20Å, aumentando significativamente o número de reflexões únicas consideradas. Na tentativa de identificar os monômeros nas regiões desordenadas como NahK ou NahL, utilizamos um modelo menos tendencioso para gerar fases iniciais. Esse modelo foi construído a partir do modelo proposto por Dr. Guimarães, trocando por alaninas todos os resíduos que diferem quando NahK e NahL são estruturalmente superpostas. A análise dos mapas de densidade eletrônica e de diferença de Fourier permitiu identificar assertivamente as cadeias laterais dos resíduos que diferenciavam as duas enzimas em 17 dos 20 monômeros que compõem a unidade assimétrica do cristal do complexo.

Os três monômeros restantes demandaram uma nova análise do mapa de densidade eletrônica experimental, dessa vez calculado para um modelo onde os três sítios correspondentes eram ocupados por cadeias de alaninas e resíduos KL. Após alguns ciclos de refinamento, um dos três monômeros pôde ser identificado por NahL. Os outros dois não foram identificados e foi proposto um modelo onde os sítios correspondentes poderiam ser ocupados tanto pela enzima NahK quanto pela enzima NahL, de maneira desordenada, formando heterodecâmeros distintos ao longo do cristal. As estatísticas dos refinamentos feitos usando esse novo modelo de desordem e os dados de XRD reprocessados com ambos limites de resolução, 47,66Å – 2,49Å e 50,0Å – 2,20Å De fato, independentemente dos dados utilizados no refinamento, foram observadas regiões com mapa de densidade eletrônica experimental pobre, o que sugere que esta é uma característica da amostra. Assim, o modelo de desordem para os monômeros indefinidos, onde não forçamos a ocupação dos sítios com densidade eletrônica pobre nem pela enzima NahK, nem por NahL parece ser o mais coerente com a densidade eletrônica observada sem deixar de ser plausível do ponto de vista bioquímico.

Capítulo 5

BPV E2

5.1 Introdução

Os papiloma vírus são um grupo de vírus de DNA circular de cadeia dupla, responsáveis tanto por lesões (verrugas) quanto por carcinomas malignos em tecidos epiteliais [109], e também estão associados a outras doenças em humanos e outros mamíferos [110], [111]. O genoma desses vírus contém aproximadamente 7900 pares de base e contém fases de leitura aberta (ORF – *open reading frame*) de transcrição precoce e de transcrição tardia. Particularmente, duas ORF's de transcrição precoce, E1 e E2, e duas de transcrição tardia, L1 e L2, são altamente conservadas dentre os membros da família dos papiloma vírus. E1 e E2 são os primeiros genes expressos durante a infecção e responsáveis pelo controle da transcrição dos genes virais e da replicação do genoma viral, enquanto L1 e L2 são proteínas estruturais que compõem a cápsula viral.[112].

Existiu um grande interesse no estudo dos vários tipos de papiloma vírus humano, que hoje já são melhor compreendidos. Nesse processo, o papiloma vírus bovino tipo 1 (BPV-1) serviu como protótipo para estudos de biologia molecular e genética desse grupo de vírus, e foi extensivamente estudado [113].

A proteína E2 do BPV-1

A proteína E2, produto da ORF E2 do BPV-1, é composta por 410 aminoácidos e sua estrutura pode ser dividida em três regiões distintas: o domínio amino-terminal, com 160 resíduos de aminoácidos, um conector, cuja estrutura

primária é rica em prolínas, e o domínio carboxi-terminal (C-terminal), com 85 resíduos de aminoácidos que se liga de forma específica ao DNA. O domínio C-terminal (E2/326-410), forma um barril- β com α -hélices na superfície e se liga ao DNA como um dímero. Sua estrutura primária é constituída dos aminoácidos:

**SCFALISGTANQVKCYRFRVKKNHRHRYENCTTTWFTVADNGAERQGQAQILITFGSPSQRQDFLKH
VPLPPGMNISGFTASLDF**

que recebem a numeração de 326 a 410. A massa molecular estimada a partir da sequência de aminoácidos é de 9,6kDa.

A interação do dímero com o DNA é mediada pelos resíduos Asn 336, Lys 339, Cys 340, Phe 343 que compõem uma das α -hélices na superfície (HI – resíduos de 335-349) [114], [115]. A estrutura tridimensional de E2/326-410 do BPV-1 foi previamente determinada por NMR (PDB ID: 1DBD) [116] e por difração de raios X por monocristal para um cristal de dímeros de E2/326-410 cristalizados no grupo de espaço P6₅22 (PDB ID: 1JJH) [115] e para um cristal de dímeros de E2/326-410 complexados com o DNA no grupo de espaço R32 (PDB ID: 2BOP) [114].

Nesse trabalho, reportamos a estrutura do domínio E2/326-410 estendido por 17 resíduos extras adicionados ao N-terminal determinada por difração de raios X. A cristalização e os experimentos de difração, assim como o trabalho inicial com os dados foi realizado por Barbosa *et al.* [117]. O solução definitiva da estrutura, as estratégias de refinamento e a validação da estrutura são detalhados nas próximas seções.

5.2 Procedimentos e discussão

Cristalização e difração de raios X

As etapas mencionadas nessa subseção foram executadas anteriormente ao período do meu mestrado e estão reportadas no trabalho de Barbosa *et al.*, 2001[117]

A clonagem, expressão e purificação da proteína E2/326-410 BPV-1 foram realizados junto ao grupo do Dr. Gonzalo Prat Gay no *Instituto de Investigaciones Bioquímicas Fundación Campomar*, Argentina. O vetor utilizado na clonagem codificava uma cauda com os 17 resíduos, MVPRRTTNDGFHLLKAG, que foi adicionada ao N-terminal da proteína. Cristais foram obtidos por difusão de vapor de gota pendurada a 18°C, onde a solução de proteína continha 8.64mg/ml de proteína, 100mM NaCl, 10mM DTT em tampão de 25mM TRIS, pH 7,5; a composição da solução de reservatório era 100mM de citrato de sódio pH 4.6-5.2, 28-32% PEG 4000 e 200mM de acetato de amônio.

A coleta de dados de difração de raios X de cristais de E2/326-410 BPV-1 com os 17 resíduos adicionais, foi realizada no Laboratório Nacional de Luz Síncrotron (LNLS), Campinas - SP. O processamento inicial dos dados foi feito para resolução com limites 30,0–2.0Å, utilizando o software HKL 2000 e permitiu a determinação dos parâmetros da célula unitária $a = b = 55,3\text{Å}$, $c = 203,4\text{Å}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, e a determinação do grupo de espaço que descreve a simetria do cristal, $P3_121$. Uma substituição molecular e o refinamento inicial da estrutura permitiu visualizar quatro monômeros formando dois dímeros funcionais na unidade assimétrica, entretanto os refinamentos não chegaram a uma etapa conclusiva.

Reprocessamento de dados

Uma nova redução de dados foi feita utilizando o software *XDS* [59], [60] que permitiu uma extensão dos limites de resolução dos dados utilizados no processamento para 40,0–1.9Å. A comparação entre as estatísticas dos processamentos é mostrada na Tabela 5.2.1, onde vemos um aumento de 82,1% no número de reflexões integradas, que corresponde a 118,9% de reflexões únicas a mais no novo processamento. O aumento considerável no número de reflexões únicas se deve ao aumento do intervalo de resolução e também à mudança metodológica de integração do *XDS* em comparação ao *HKL2000*. Os índices de qualidade do processamento, apresentados na Tabela 5.2.1, mostram que as novas reflexões adicionadas ao conjunto de dados de XRD não prejudicou

as estatísticas do processamento. Em particular, a diminuição do fator $\langle I/\sigma(I) \rangle$ é esperada, uma vez que consideramos um alto número de reflexões pouco intensas.

Tabela 5.2.1 – Estatísticas dos processamentos de dados utilizando o software *HKL2000* com limites de resolução 30,0-2,0Å, e o software *XDS*, com os limites de resolução 40,0-1,9Å.

	HKL 2000	XDS
Resolução (Å)	30,0-2,0	40,0-1,90
Reflexões	127.541	232.219
Reflexões únicas	24.850	29.354
$R_{\text{merge/meas}}(\%)*$	8,7 (62,0)	12,6 (91,5)
$\langle I/\sigma(I) \rangle$	16,7 (3,1)	10,2 (1,8)
Completeza (%)	99,8 (98,9)	99,8 (99,6)

Os valores entre parênteses se referem à última camada de resolução.

*Para os dados processados com o software *HKL2000*, temos R_{merge} , e para dados processados com o software *XDS*, temos R_{meas} .

Solução da estrutura e refinamento

A partir dos dados reprocessados, uma nova substituição molecular utilizando o software *PHASER* [63] da plataforma CCP4 [62] foi realizada. O modelo estrutural utilizado foi a estrutura de E2/326-410 de BPV-1 complexada com DNA (PDB ID: 2BOP) e uma única solução foi encontrada, na qual quatro monômeros de E2/326-410 ocupam a unidade assimétrica arranjados dois a dois em dois barris- β com α -hélices na superfície, concordando com o modelo proposto por Barbosa et al.. Os quatro monômeros arranjados em dois dímeros funcionais podem ser visualizados na Figura 5.2.1.

Ciclos de refinamento foram realizados de forma exaustiva, utilizando o software *REFMAC5* [65], acompanhados da inspeção visual de todas as cadeias superpostas ao mapa de densidade eletrônica calculado, utilizando o software

COOT [66]. Os dados de XRD permitiram a construção de um mapa de densidade eletrônica preciso ao longo da maior parte dos monômeros e as modificações no modelo, como ajustes da geometria e do posicionamento dos aminoácidos e adição de moléculas de água, foram feitas manualmente. As estatísticas do refinamento estão mostradas na Tabela 5.2.2.

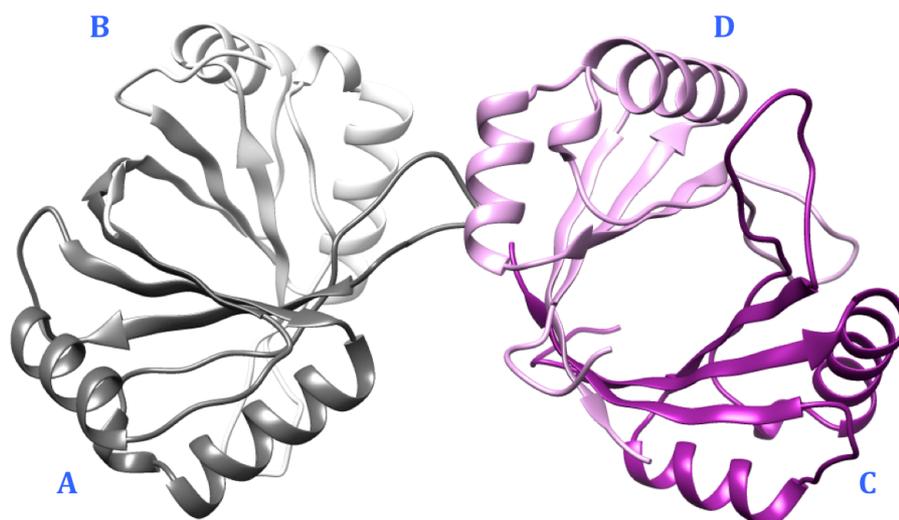


Figura 5.2.1 – Dois dímeros funcionais de E2/326-410. Os dímeros funcionais são formados pela junção dos monômeros (A+B e C+D). O barril- β é evidenciado no dímero formado pelas cadeias A e B e/ou C e D.

Tabela 5.2.2 – Estatística final do refinamento e do modelo de E2/326-410 com 17 resíduos adicionados ao N-terminal.

Resolução (Å)	40,0-1,90 (2,02-1,90)
$R_{\text{factor}}/R_{\text{free}}$	0,192/0,246
Desvio RMS	
ligações	0,016
ângulos	1,766

Fator B médio	
cadeia A	27,8
cadeia B	27,6
cadeia C	29,1
cadeia D	31,5
Número de resíduos	371
Número de átomos	2922
Moléculas de água	199
Solvente na célula unitária (%)	44,6

Análise do modelo

Dois dímeros funcionais formados por quatro monômeros de E2/326-410, 371 resíduos de aminoácidos e 199 moléculas de água foram modelados a partir do mapa de densidade eletrônica experimental. Os quatro monômeros conservam a estrutura secundária e o enovelamento proteico. A superposição estrutural desses monômeros é mostrada na Figura 5.2.2 onde duas regiões, que correspondem ao loop (alça) entre os resíduos 365 e 371 (loop I), inclusive, e ao N-terminal (extremidade desordenada), se destacam pela diferença em suas posições.

Essas duas regiões apresentaram um ajuste pobre do modelo à densidade eletrônica. Exceções são o ajuste do N-terminal da cadeia C e do loop da cadeia A, onde o mapa de densidade eletrônica calculado era mais preciso. Dos 17 resíduos adicionados ao N-terminal, apenas 6 foram observados em todos os quatro monômeros na unidade assimétrica. Particularmente, no monômero C, 9

desses resíduos puderam ser modelados. Seu N-terminal mais longo é mostrado na Figura 5.2.2. Também foi feita a modelagem completa do loop I na cadeia A.

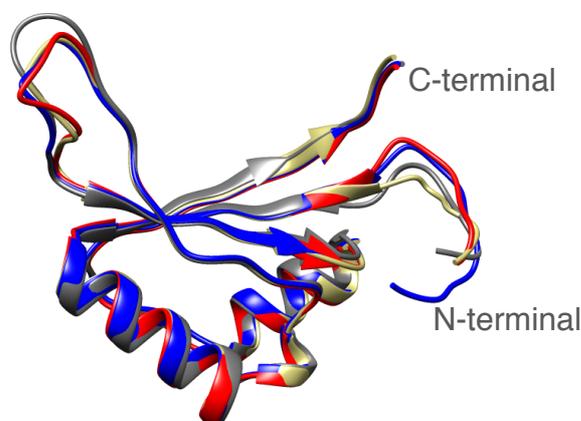


Figura 5.2.2 – Superposição estrutural dos quatro monômeros presentes na unidade assimétrica do modelo: cadeia A (vermelho), B (amarelo), C (azul) e D (cinza); as porções N e C-terminal dos monômeros estão indicadas por N e C respectivamente.

Os fatores de deslocamento isotrópico (seção 2.4) – chamados de Fator B na comunidade de cristalografia de proteína - dos quatro monômeros estão mostrados na Figura 5.2.3. Valores elevados refletem as regiões onde o ajuste do modelo ao mapa é ruim. Novamente destacam-se as regiões correspondentes ao N-terminal e ao loop I. Utilizando-se o algoritmo *MATTHEWS_COEF* [118], [119], também da plataforma CCP4, calculou-se uma estimativa de conteúdo de solvente na célula unitária, que indicou que 44,6% de seu volume era preenchido por água. Assim como a maioria dos empacotamentos cristalinos proteicos, a porcentagem de solvente é grande [118], permitindo a mobilidade de algumas regiões da proteína de tal forma que os elétrons nessas regiões, que são os espalhadores de raios X, não possuem uma posição única ao longo do cristal. Como detalhado na seção 2.4, a função densidade eletrônica (Equação 2.4. 37) depende diretamente da posição dos espalhadores no espaço real, e sempre que muita mobilidade é conferida a uma parte da estrutura, a densidade eletrônica calculada correspondente é pouco precisa ou até não observada. O N-terminal e a alça se localizam em regiões com alta exposição ao solvente, o que pode ter garantido a eles a alta mobilidade que justifica os altos fatores B e o mapa de densidade eletrônica pouco preciso.

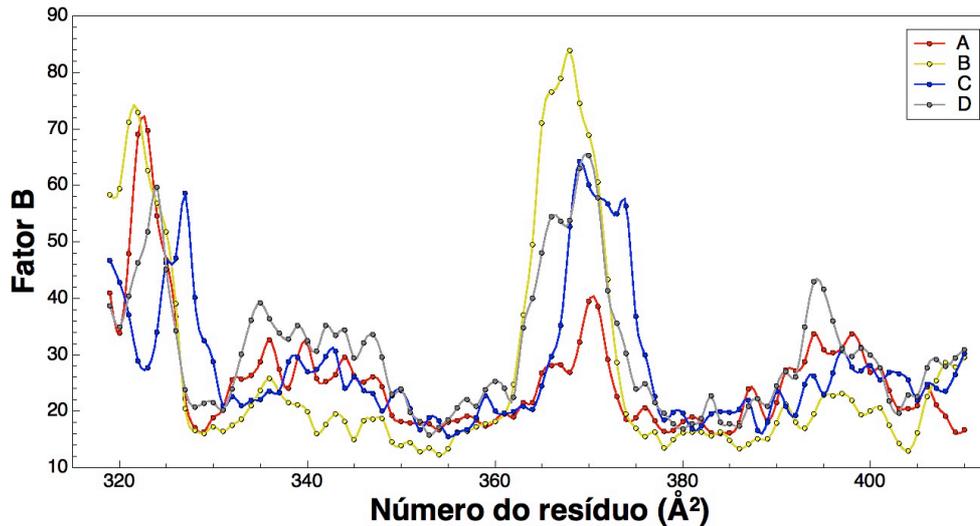


Figura 5.2.3 – Fatores B para as quatro cadeias modeladas nesse trabalho. Linhas são guias para os olhos.

Todos os resíduos foram ajustados para obedecer as restrições dos ângulos ψ e ϕ (ver seção 2.1). O gráfico de Ramachandran na Figura 5.2.4 mostra que todos os resíduos possuem configurações aceitáveis: 12 deles apresentam configurações permitidas, ou seja, menos favoráveis e todos os demais ocupam as regiões favoráveis.

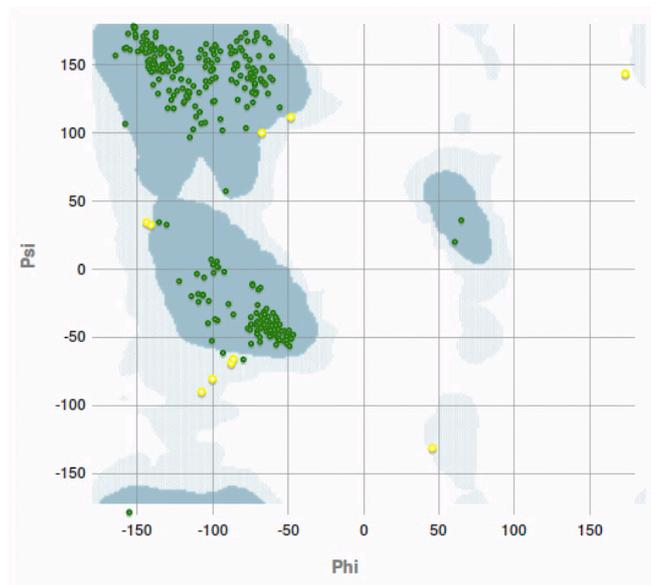


Figura 5.2.4 – Gráfico de Ramachandran para o modelo com dois dímeros funcionais de E2/326-410. Resíduos representados em verde estão em regiões favoráveis e resíduos em amarelo estão em regiões permitidas (= menos favoráveis). Nenhum resíduo apresenta configuração proibida.

Comparação com as estruturas do PDB

A estrutura depositada no PDB com ID 1JJH foi determinada por difração de raios X a 2,5Å de resolução [115]. Ela contém dímeros de E2/326-410 conectados por pontes dissulfeto entre as cisteínas 340 (Cys340) como mostrado na Figura 5.2.5. Nesse modelo, o loop I entre os resíduos 365 e 371 não foi observado. Já a estrutura depositada no PDB com ID 2BOP foi determinada por difração de raios X a 1,7Å de resolução [114]. Ela contém dímeros de E2/326-410 complexados com o DNA, como mostrado na Figura 5.2.9. A interação com o DNA é mediada unicamente pelos aminoácidos Asn 336, Lys 339, Cys 340 e Phe 343 pertencentes à hélice de reconhecimento (entre os resíduos 335 e 349).

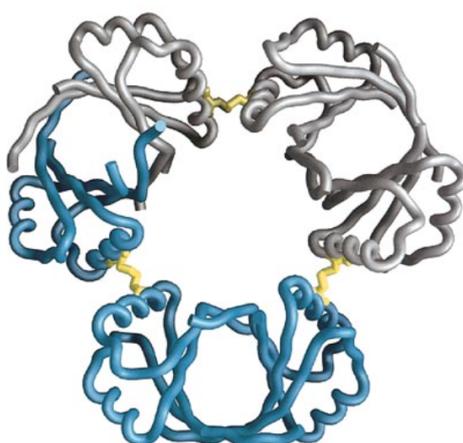


Figura 5.2.5 – Arranjo dos dímeros ao longo da rede cristalina na estrutura 1JJH. As pontes dissulfeto entre as Cys340 estão indicadas em amarelo. O hexâmetro é formado pelos monômeros pertencentes a duas unidades assimétricas adjacentes, mostrados em azul e cinza.

Como mencionado anteriormente, no modelo com dois dímeros funcionais proposto nesse trabalho, o loop I pôde ser completamente modelado para a cadeia A, ou seja, pela primeira vez ele foi observado em uma estrutura resolvida por difração de raios X por monocristal de E2/326-410 não complexada com o DNA. Seu ajuste à densidade eletrônica é mostrado na Figura 5.2.6. A disposição dos dímeros, mostrada na Figura 5.2.1, fez com que o loop I da cadeia A ficasse especialmente localizado próximo de uma das α -hélice da cadeia D, o que possivelmente restringiu sua mobilidade no cristal e permitiu sua modelagem. De farto, os dímeros não se conectam através de pontes dissulfeto como em 1JJH,

mas interagem fracamente. A Cys 340, envolvida no empacotamento cristalino de 1JJH, é um dos resíduos da que interage especificamente com o DNA e, uma vez que em 1JJH ela participa de uma interação específica da formação do cristal, o contraste entre sua conformação quando E2/326-410 interage ou não com o DNA ainda não havia sido possível.

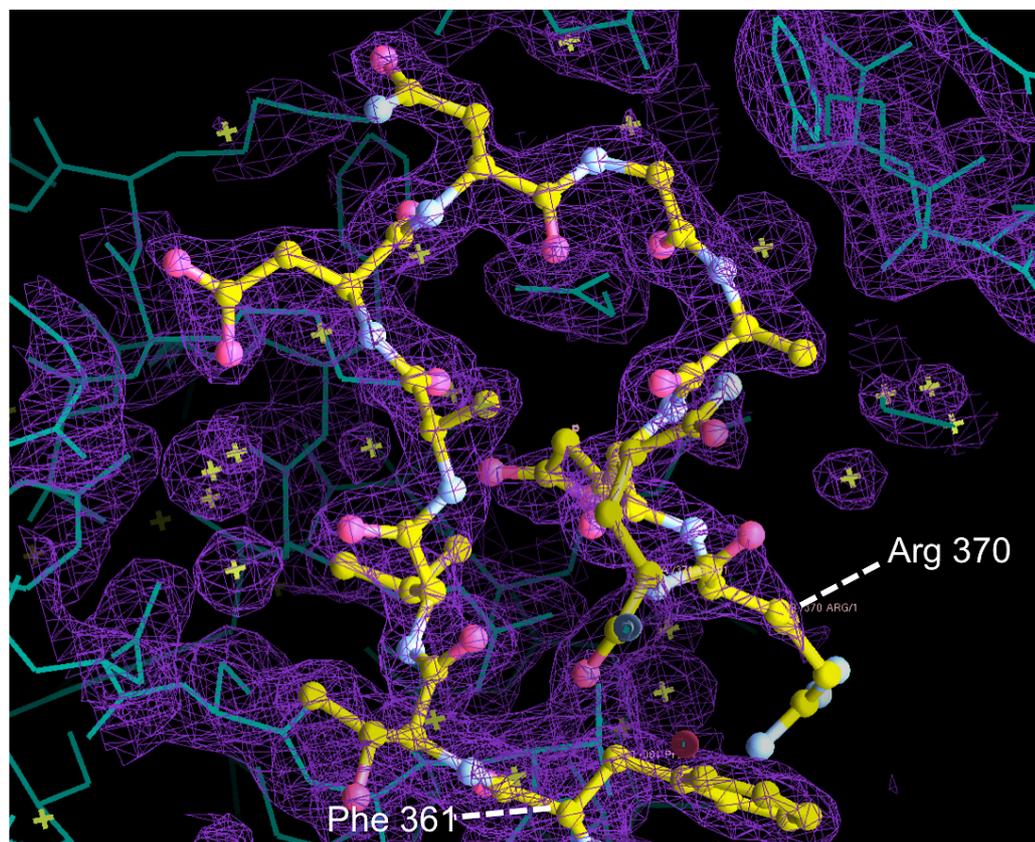


Figura 5.2.6 – Ajuste do loop I (amarelo) ao mapa de densidade eletrônica (roxo) com 0.46 elétron por \AA^3 . Os resíduos fenilalanina 261 (Phe 261) e arginina 370 (Arg 370) estão destacados.

Foi feita a superposição de todas as hélices de reconhecimento de DNA de E2/326-410: dos quatro monômeros modelados nesse trabalho, das duas cadeias de 1JJH e do monômero de 2BOP¹². A mudança conformacional da Cys340 em função da ligação, seja com o DNA em 2BOP, seja com as outras cisteínas em 1JJH, é mostrada na Figura 5.2.7. Vê-se que a cadeia lateral das cisteínas apresenta uma pequena mudança no ângulo formado pelas ligações C-C-S, que é maior para as Cys livres. Contudo não foi observada nenhuma outra

¹² Em 1JJH, o terceiro monômero na unidade assimétrica é gerado por operações de simetria não cristalográfica, assim como o dímero em 2BOP é gerado a partir de um monômero apenas.

mudança estrutural que permitisse inferir qualquer influência da ligação de Cys340 na estrutura.

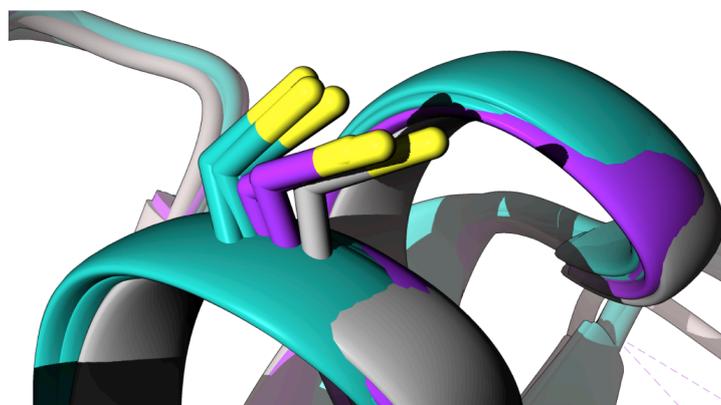


Figura 5.2.7 - Cisteína 340 em 2BOP (cinza), nos dois monômeros de 1jjH (roxo) e nos quatro monômeros modelados nesse tranalho (ciano).

Por fim, a superposição da cadeia A com um monômero de 2BOP é mostrada na Figura 5.2.9. Vemos a superposição quase total das estruturas O loop I destaca-se com alto RMSD (Figura 5.2.8), entretanto não conseguimos determinar se a grande mudança na posição da alça se deve ao empacotamento da estrutura, à ligação do dímero com o DNA ou ambos.

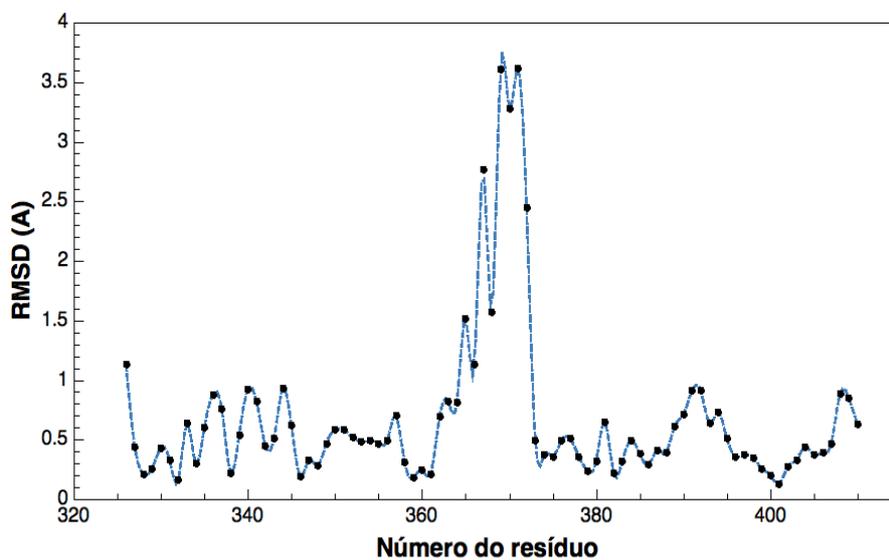


Figura 5.2.8 - Gráfico do desvio quadrático médio das posições das cadeias principais de 2BOP e da cadeia A.

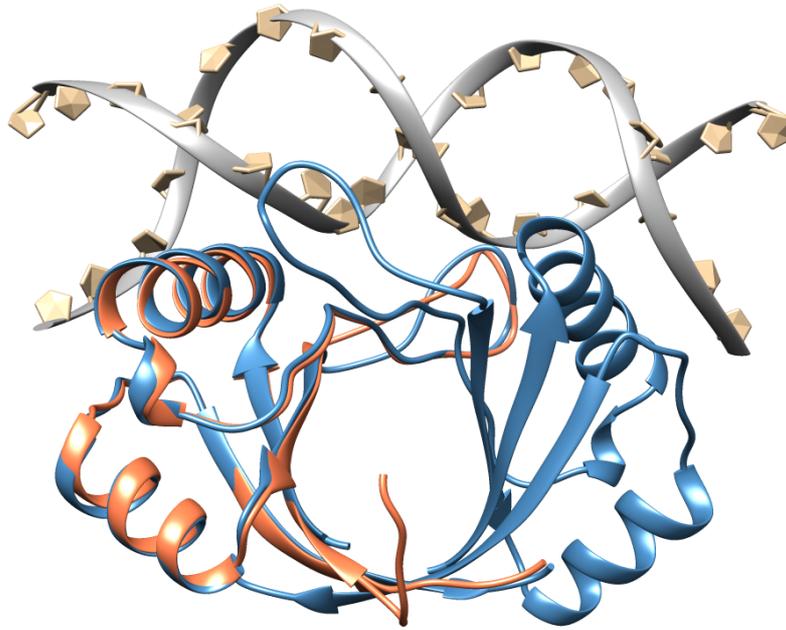


Figura 5.2.9 – Cadeia A do dímero funcional (laranja) superposto a estrutura previamente determinada (PDB ID 2BOP) de E2/326-410 (azul) complexada com o DNA (cinza).

5.3 Conclusão

Os papiloma vírus infectam células epiteliais e ainda são associados a outras doenças em humanos e outros mamíferos. As proteínas que atuam como fatores de virulência desse grupo de vírus foram extensivamente estudadas, processo no qual o papiloma vírus bovino serviu como protótipo.

A proteína E2/326-410 do papiloma vírus bovino tipo 1 com 17 resíduos extras adicionados ao N-terminal foi cristalizada no grupo de espaço $P3_121$. Dados obtidos em experimentos de difração de raios X indicaram a presença de dois dímeros funcionais na unidade assimétrica, como reportado por Barbosa *et al.* [117]. A estrutura de E2/326-410 havia sido previamente determinada por difração de raios X de monocristal para a proteína cristalizada nos grupos de espaço $P6_522$ (PDB ID: 1JJH) e $R32$ (PDB ID: 2BOP), sendo o último a estrutura de E2/326-410 complexada com o DNA.

Nesse trabalho, foi feito o reprocessamento dos dados de XRD, uma nova substituição molecular e uma série de etapas de refinamento, utilizando softwares desenvolvidos mais recentemente. A estrutura dos dois dímeros foi devidamente elucidada e um modelo estrutural foi proposto. A interação dos dímeros nessa configuração cristalina não foi mediada por uma ponte dissulfeto, como previamente observado em 1JJH, e reduziu a mobilidade de um dos loops na estrutura permitindo a sua modelagem, o que não havia sido feito anteriormente para E2/326-410 não complexada com o DNA. Dos 17 resíduos adicionados ao N-terminal, apenas 6 foram observados em todas as cadeias e especialmente a cadeia A teve 9 desses resíduos modelados.

Uma comparação com a estruturas 1JJH e 2BOP mostrou uma pequena variação na conformação da cadeia lateral de Cys340, resíduo que interage especificamente com o DNA. No geral, o enovelamento proteico em solução ou no ambiente cristalino se mantém. Mas também o empacotamento cristalino pode levar à pequenas mudanças conformacionais, principalmente em loops e cadeias laterais [120], o que foi definitivamente confirmado para as Cys340 que se conectam covalentemente em 1JJH. Todavia, a diferença entre a posição da cadeia lateral de Cys340 nos dímeros modelados nesse trabalho e em 2BOP é pequena e nenhuma outra mudança estrutural que permitisse inferir qualquer influência estrutural da ligação de Cys340 foi observada.

Por fim, a superposição da cadeia A de um dos dímeros com um monômero de 2BOP mostrou que o loop I se destaca com alto RMSD. Não conseguimos determinar se a grande mudança na posição da alça se deve ao empacotamento dos dímeros funcionais, à ligação do dímero com o DNA ou ambos. A indeterminação na posição dos resíduos que compõe o loop I, e suas respectivas cadeias laterais, nas cadeias B, C e D e nos monômeros de 1JJH impossibilitou comparações adicionais.

Conclusão e perspectivas

As proteínas são responsáveis por inúmeros processos biológicos vitais. A função e a forma dessas moléculas estão intimamente relacionadas, logo, o conhecimento da estrutura tridimensional de proteínas é essencial para a descrição dos mecanismos que elas mediam. A cristalografia é uma poderosa técnica empregada na determinação estrutural de proteínas e sua aplicação envolve várias etapas, cuja execução exige vasto conhecimento biologia molecular, bioquímica e física.

Nesse trabalho visitamos todas as etapas envolvidas no processo de determinação da estrutura tridimensional de proteínas por cristalografia. Três proteínas foram estudadas: uma esfingomielinase D (SMase D), a proteína E2 do papiloma vírus bovino tipo 1 (E2 de BPV-1) e o complexo das enzimas NahK/NahL.

A SMase D de *C. ulcerans* foi clonada em pET28a-TEV, expressa em *E. coli*, e, após purificada, apresentou monodispersa em ensaios de DLS. A análise do espectro de dicroísmo circular indicou uma alta porcentagem (>90%) de α -hélices na estrutura da enzima e mostrou que a proteína mantém parte de sua estrutura secundária até $\sim 50^\circ\text{C}$. A modelagem partindo da estrutura primária sugeriu um barril-TIM característico de proteínas da família. Cristais macroscópicos foram obtidos em ensaios no RoboLab para três condições distintas, entretanto eles não difrataram quando submetidos a experimentos de difração de raios X (XRD). À vista disso, devem ser então realizadas outras tentativas de otimização das condições de cristalização encontradas a fim de se obter cristais mais adequados para experimentos de XRD.

Dados de XRD obtidos para um cristal do complexo NahK/NahL de *P. putida* foram reprocessados com limites de resolução estendidos em busca de mais informações que permitissem a solução da desordem observada no mapa de densidade eletrônica experimental calculado. Foi proposto um novo modelo estrutural no qual as enzimas NahK e NahL ocupavam de forma desordenada e simultaneamente um sítio de cada heterodecâmero na unidade assimétrica. Esse modelo foi refinado e o parâmetro da ocupação de cada enzima convergiu para valores próximos de 0,5 em todos os sítios. Esse novo modelo é mais coerente com a densidade eletrônica observada, que sugere que a desordem é uma característica intrínseca da amostra. Ele deverá agora passar por ajustes finais e será discutido em uma artigo, onde o mecanismo de ação dessas enzimas será descrito.

Por fim, dados de XRD de um cristal da proteína E2 de BPV-1, cuja simetria não havia sido observada anteriormente, foram reprocessados e um modelo estrutural foi proposto e validado. Nesse modelo, a interação entre os dímeros na unidade assimétrica reduziu a mobilidade de um dos loops da proteína e permitiu sua modelagem, que não havia sido reportada para E2 não complexada com o DNA. Também foi possível avaliar a mudança conformacional do resíduo Cys340 quando esse se liga covalentemente a outra Cys340, interage especificamente com o DNA e quando não participa de nenhuma interação intermolecular dentro do cristal. Observamos uma pequena mudança na posição da cadeia lateral de Cys 340, que parece ser apenas devida ao empacotamento cristalino. O novo modelo estrutural da proteína E2 de BPV-1 e sua comparação com os modelos já publicados serão discutidos em um do artigo a ser submetido.

O estudo das três proteínas constituiu um longo processo de aprendizado que envolveu o domínio de inúmeras técnicas experimentais e a utilização de vários softwares dedicados à cristalografia de proteínas. E, de fato, muitas horas de trabalho nos laboratórios de cristalografia (LabCri) e de biologia estrutural (BioEst) permitiram uma compreensão geral do processo de determinação da estrutura tridimensional de proteínas por técnicas de difração de raios X.

Apêndices

Apêndice A Proteínas

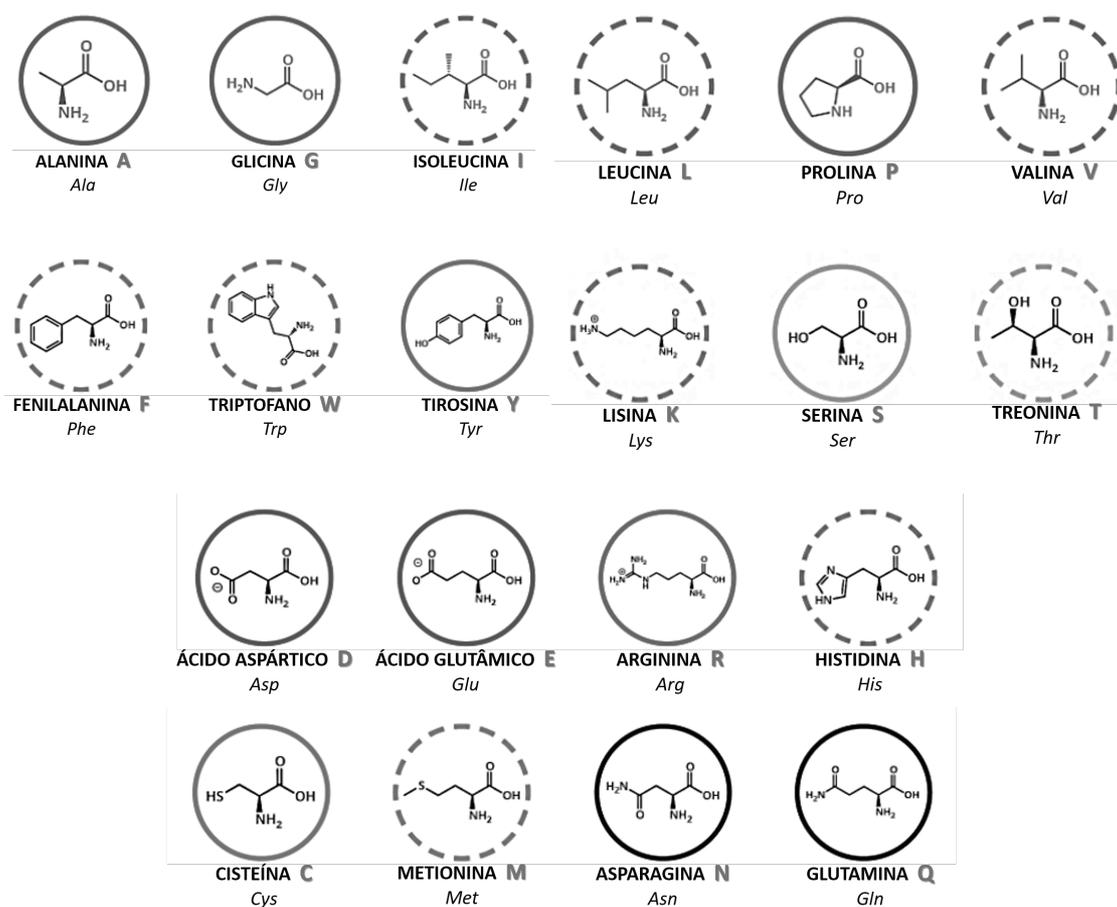


Figura 6.1 – Os 20 L-aminoácidos mais comuns em proteínas. Fonte [121]

Tabela 6.1 – Sequência de bases, códons, e respectivos aminoácidos transcritos. Importante notar que muitas vezes a variação da terceira base é irrelevante para a definição do aminoácido. Os códons que indicam o ponto final da proteína estão destacados. Para mais detalhes sobre códons de iniciação e parada, ver [1].

1ª base	2ª base				3ª base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	<i>Parada</i>	<i>Parada</i>	A
	Leu	Ser	<i>Parada</i>	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Ans	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

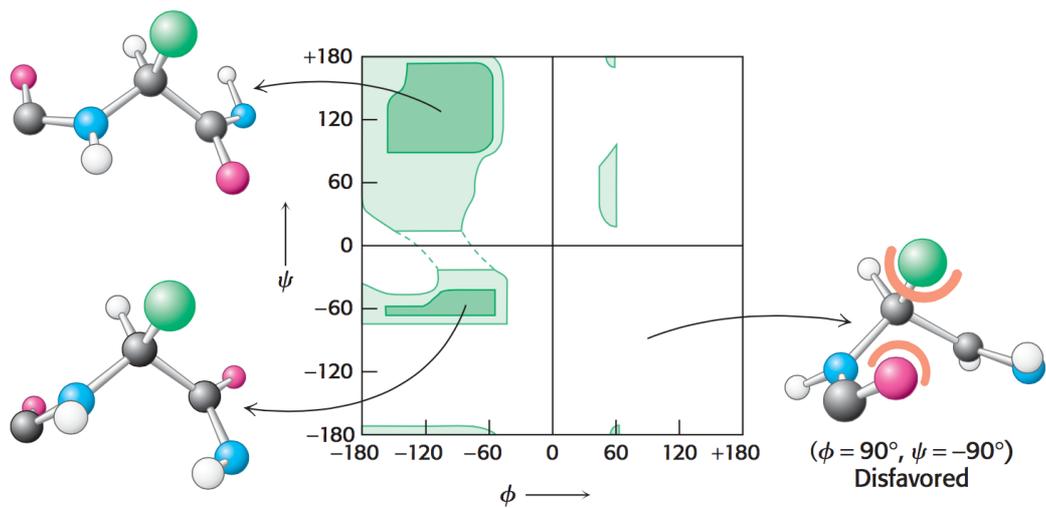


Figura 6.2 – Gráfico de Ramachandran, mostrando os valores de ϕ e ψ . Nem todos os valores de ϕ e ψ são possíveis sem que haja colisão entre os átomos. As regiões mais favoráveis estão mostradas em verde escuro, e as regiões aceitáveis em verde claro. As duas estruturas à esquerda mostram exemplos da posição dos átomos em duas regiões permitidas. À direita vemos uma conformação proibida impedimento estérico (*disfavored*, do inglês, defavorecido). Fonte: [1]

Apêndice B– Complexo NahKL

Tabela 6.2 – Resultados da inspeção visual da densidade eletrônica nos sítios G, Q e R do complexo NahK/NahL. Os resíduos analisados se superpõem no alinhamento estrutural das enzimas e apresentam alta dissimilaridade. Os resíduos cuja densidade eletrônica correspondente foi observada estão identificados por um X.

		G		Q		R	
NahK	NahL	NahK	NahL	NahK	NahL	NahK	NahL
ALA 12	GLU 8	-	-	-	-	-	-
HIS 16	GLU 12	-	-	-	-	-	-
GLU 18	TYR 14	-	-	-	-	-	-
GLU 21	MET 17		X	-	-	-	-
VAL 24	ARG 20	-	-	-	-	-	-
HIS 25	GLU 21	-	-	-	-	-	-
ASP 26	ALA 22	-	-	-	-	-	-
LYS 29	PRO 25	-	-	-	-	-	-
GLU 36	ALA 32	-	-	-	-	-	-
ALA 40	ASP 36	-	-	-	-	-	-
ASP 44	HIS 40	-	-	-	-	X	
GLN 46	SER 42	-	-	-	-	-	-
ARG 50	LEU 46	-	-	-	-	-	-
TRP 70	LYS 66	-	-	-	-	-	-
MET 73	GLN 69	X		-	-	-	-
GLU 79	TYR 75	-	-	-	-	-	-
PRO 93	ASN 89	X		-	-		X
GLY 96	GLU 92		X	-	-	-	-
ASP 99	PRO 95	-	-	-	-	-	-
CYS 100	SER 97	-	-	-	-	-	-

HIS 105	GLN 102	-	-	-
THR 116	LEU 113	-	-	-
ALA 118	LYS 115	-	-	-
PRO 119	ASP 116	-	-	-
HIS 121	GLY 119	-	-	-
HIS 126	THR 123	X	X	X
TYR 136	CYS 133	X	X	X
THR 140	CYS 137	-	-	-
VAL 141	PHE 138	-	-	-
ASN 150	TRP 148	-	-	-
PHE 153	ILE 150	-	-	-
THR 166	GLY 163	-	-	-
ARG 167	LEU 164	X	X	-
GLY 172	ASP 169	X	-	-
MET 174	ALA 171	X	-	-
LEU 177	PRO 174	-	X	-
MET 189	VAL 186	-	-	-
GLU 197	SER 194	-	-	-
HIS 207	SER 204	X	X	-
MET 214	TRP 211	-	-	-
ARG 222	PHE 219	X	-	-
HIS 225	PRO 222	X	X	X
PHE 231	VAL 228	-	-	-
ASN 248	PHE 245	-	X	-
THR 250	ARG 247	-	X	X
TYR 253	ILE 250	-	-	-
GLN 254	GLY 251	X	X	-
VAL 187	MET 184	X	X	X

MET 37	ILE33	-	-	-	-	-
MET 65	ILE 61	X	-	X		
GLN 75	MET 71	X	X	-		
MET 76	LEU 72	-	-	-		
ILE 104	MET 101	-	X	-		
ILE 138	MET 135	X	X	X		
VAL 187	MET 184	X	X	X		
MET 233	LEU 230	X	X	-		
ILE 249	MET 246	X	X	X		
ASP 33	ARG 29	-	-	-		
PHE 90	MET 86	-	-	-		
TOTAL:		9	11	3	13	4
						7

Referências

- [1] J. M. Berg, J. L. Tymoczko, L. Stryer, e N. D. Clarke, “Biochemistry. 2002”, *N. Y. N. Y.*, vol. 10010, 2002.
- [2] V. Ramakrishnan, “Ribosome structure and the mechanism of translation”, *Cell*, vol. 108, nº 4, p. 557–572, 2002.
- [3] H. M. Berman *et al.*, “The Protein Data Bank”, *Nucleic Acids Res.*, vol. 28, nº 1, p. 235–242, jan. 2000.
- [4] I. Russo Krauss, A. Merlino, A. Vergara, e F. Sica, “An overview of biological macromolecule crystallization”, *Int. J. Mol. Sci.*, vol. 14, nº 6, p. 11643–11691, 2013.
- [5] A. McPherson e J. A. Gavira, “Introduction to protein crystallization”, *Acta Crystallogr. Sect. F Struct. Biol. Commun.*, vol. 70, nº 1, p. 2–20, jan. 2014.
- [6] M. Karplus e J. Kuriyan, “Molecular dynamics and protein function”, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, nº 19, p. 6679–6685, maio 2005.
- [7] D. Eisenberg, E. M. Marcotte, I. Xenarios, e T. O. Yeates, “Protein function in the post-genomic era”, *Nature*, vol. 405, nº 6788, p. 823–826, jun. 2000.
- [8] J. D. Watson e F. H. Crick, “The structure of DNA”, in *Cold Spring Harbor symposia on quantitative biology*, 1953, vol. 18, p. 123–131.
- [9] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, e D. C. Phillips, “A three-dimensional model of the myoglobin molecule obtained by x-ray analysis”, *Nature*, vol. 181, nº 4610, p. 662–666, 1958.
- [10] I. Sillitoe *et al.*, “CATH: comprehensive structural and functional annotations for genome sequences”, *Nucleic Acids Res.*, vol. 43, nº D1, p. D376–D381, 2015.
- [11] S. D. Lam *et al.*, “Gene3D: expanding the utility of domain assignments”, *Nucleic Acids Res.*, vol. 44, nº D1, p. D404–D409, 2016.
- [12] M.-J. Gething e J. Sambrook, “Protein folding in the cell”, *Nature*, vol. 355, nº 6355, p. 33, 1992.
- [13] C. M. Dobson, “Protein folding and misfolding”, *Nature*, vol. 426, nº 6968, p. 884–890, 2003.

- [14] C. M. Dobson, A. Šali, e M. Karplus, "Protein folding: a perspective from theory and experiment", *Angew. Chem. Int. Ed.*, vol. 37, n° 7, p. 868–893, 1998.
- [15] D. L. Nelson, A. L. Lehninger, e M. M. Cox, *Lehninger principles of biochemistry*. Macmillan, 2008.
- [16] J. J. Pasternak, "Recombinant DNA Technology", *Introd. Hum. Mol. Genet. Mech. Inherit. Dis. Second Ed.*, p. 107–152, 2005.
- [17] T. A. Brown, *Gene cloning and DNA analysis: an introduction*. John Wiley & Sons, 2016.
- [18] D. Sherwood e J. Cooper, *Crystals, x-rays and proteins: comprehensive protein crystallography*. OUP Oxford, 2010.
- [19] J. Arnau, C. Lauritzen, G. E. Petersen, e J. Pedersen, "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins", *Protein Expr. Purif.*, vol. 48, n° 1, p. 1–13, 2006.
- [20] H. Edelhoch, "Spectroscopic determination of tryptophan and tyrosine in proteins", *Biochemistry (Mosc.)*, vol. 6, n° 7, p. 1948–1954, 1967.
- [21] A. McPherson e others, *Crystallization of biological macromolecules*. Cold Spring Harbor Laboratory Press, 1999.
- [22] H. Determann e I. Walter, "Source of aromatic affinity to 'Sephadex' dextran gels", *Nature*, vol. 219, n° 5154, p. 604–605, 1968.
- [23] C. Mateo *et al.*, "Affinity chromatography of polyhistidine tagged enzymes: New dextran-coated immobilized metal ion affinity chromatography matrices for prevention of undesired multipoint adsorptions", *J. Chromatogr. A*, vol. 915, n° 1, p. 97–106, 2001.
- [24] A. R. Ferré-D'Amaré e S. K. Burley, "Dynamic light scattering in evaluating crystallizability of macromolecules", *Methods Enzymol.*, vol. 276, p. 157–166, 1997.
- [25] A. Micsonai *et al.*, "Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy", *Proc. Natl. Acad. Sci.*, vol. 112, n° 24, p. E3095–E3103, 2015.
- [26] B. J. Berne e R. Pecora, *Dynamic light scattering: with applications to chemistry, biology, and physics*. Courier Corporation, 2000.
- [27] J. Lim, S. P. Yeap, H. X. Che, e S. C. Low, "Characterization of magnetic nanoparticle by dynamic light scattering", *Nanoscale Res. Lett.*, vol. 8, n° 1, p. 381, 2013.
- [28] S. M. Kelly, T. J. Jess, e N. C. Price, "How to study proteins by circular dichroism", *Biochim. Biophys. Acta BBA-Proteins Proteomics*, vol. 1751, n° 2, p. 119–139, 2005.

- [29] M. D. Bazzi e R. W. Woody, "Oriented secondary structure in integral membrane proteins. I. Circular dichroism and infrared spectroscopy of cytochrome oxidase in multilamellar films", *Biophys. J.*, vol. 48, n° 6, p. 957–966, 1985.
- [30] W. C. Johnson, "Protein secondary structure and circular dichroism: a practical guide", *Proteins Struct. Funct. Bioinforma.*, vol. 7, n° 3, p. 205–214, 1990.
- [31] G. Böhm, R. Muhr, e R. Jaenicke, "Quantitative analysis of protein far UV circular dichroism spectra by neural networks", *Protein Eng.*, vol. 5, n° 3, p. 191–195, 1992.
- [32] N. Sreerama e R. W. Woody, "A self-consistent method for the analysis of protein secondary structure from circular dichroism", *Anal. Biochem.*, vol. 209, n° 1, p. 32–44, 1993.
- [33] S. W. Provencher, "CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations", *Comput. Phys. Commun.*, vol. 27, n° 3, p. 229–242, 1982.
- [34] N. Sreerama e R. W. Woody, "Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set", *Anal. Biochem.*, vol. 287, n° 2, p. 252–260, 2000.
- [35] "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions: Abstract: Nature Protocols". [Online]. Disponível em: <http://www.nature.com/nprot/journal/v1/n6/abs/nprot.2006.204.html>. [Acessado: 27-abr-2017].
- [36] A. Rodger e B. Nordén, *Circular Dichroism and Linear Dichroism*. Oxford University Press, 1997.
- [37] M. P. Marder, *Condensed matter physics*. John Wiley & Sons, 2010.
- [38] J. C. Brice e P. Rudolph, "Crystal Growth", *Ullmanns Encycl. Ind. Chem.*, 2007.
- [39] G. Feher e Z. Kam, "Nucleation and growth of protein crystals: General principles and assays", *Methods Enzymol.*, vol. 114, p. 77–112, 1985.
- [40] R. P. Sear, "Nucleation: theory and applications to protein solutions and colloidal suspensions", *J. Phys. Condens. Matter*, vol. 19, n° 3, p. 33101, jan. 2007.
- [41] H. Vehkamäki, *Classical nucleation theory in multicomponent systems*. Springer Science & Business Media, 2006.

- [42] J. P. Chesick, "Physical chemistry, second edition (Castellan, Gilbert W.)", 01-ago-1972. [Online]. Disponível em: <http://pubs.acs.org/doi/full/10.1021/ed049pA433.1>. [Acessado: 20-jun-2017].
- [43] A. McPherson, *Preparation and analysis of protein crystals*. John Wiley & Sons, 1982.
- [44] A. Ducruix e R. Giegé, *Crystallization of nucleic acids and proteins*. IRL Press at Oxford University Press, 1992.
- [45] J. Jancarik e S.-H. Kim, "Sparse matrix sampling: a screening method for crystallization of proteins", *J. Appl. Crystallogr.*, vol. 24, nº 4, p. 409–411, 1991.
- [46] R. Page *et al.*, "Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 59, nº 6, p. 1028–1037, jun. 2003.
- [47] G. R. Fowles, *Introduction to modern optics*. Courier Corporation, 1975.
- [48] IUCr, "Report of the Executive Committee for 1991", *Acta Crystallogr. A*, vol. 48, nº 6, p. 922–946, nov. 1992.
- [49] B. E. Warren, *X-ray Diffraction*. Courier Corporation, 1969.
- [50] A. J. Blake e W. Clegg, *Crystal structure analysis: principles and practice*, vol. 13. Oxford University Press, 2009.
- [51] M. G. Rossmann e D. M. Blow, "The detection of sub-units within the crystallographic asymmetric unit", *Acta Crystallogr.*, vol. 15, nº 1, p. 24–31, 1962.
- [52] P. T. Beurskens, R. O. Gould, H. J. Bruins Slot, e W. P. Bosman, "Translation functions for the positioning of a well oriented molecular fragment", *Z. Für Krist.-Cryst. Mater.*, vol. 179, nº 1–4, p. 127–160, 1987.
- [53] R. J. Read e A. J. Schierbeek, "A phased translation function", *J. Appl. Crystallogr.*, vol. 21, nº 5, p. 490–495, 1988.
- [54] J. Navaza e E. Vernoslova, "On the fast translation functions for molecular replacement", *Acta Crystallogr. A*, vol. 51, nº 4, p. 445–449, 1995.
- [55] A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, e R. J. Read, "Likelihood-enhanced fast translation functions", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 61, nº 4, p. 458–464, abr. 2005.
- [56] P. Evans e A. McCoy, "An introduction to molecular replacement", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 64, nº 1, p. 1–10, 2008.
- [57] C. Chothia e A. M. Lesk, "The relation between the divergence of sequence and structure in proteins.", *EMBO J.*, vol. 5, nº 4, p. 823, 1986.

- [58] Z. Otwinowski, "Oscillation data reduction program", in *Proceedings of the CCP4 Study Weekend*, 1993, p. 56–62.
- [59] W. Kabsch, "Xds", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, n° 2, p. 125–132, 2010.
- [60] W. Kabsch, "Integration, scaling, space-group assignment and post-refinement", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, n° 2, p. 133–144, 2010.
- [61] M. Gilski, "Processing and analysis of synchrotron diffraction images of protein crystals".
- [62] M. D. Winn *et al.*, "Overview of the CCP4 suite and current developments", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 67, n° 4, p. 235–242, 2011.
- [63] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, e R. J. Read, "Phaser crystallographic software", *J. Appl. Crystallogr.*, vol. 40, n° 4, p. 658–674, 2007.
- [64] E. Krissinel, "Enhanced fold recognition using efficient short fragment clustering", *J. Mol. Biochem.*, vol. 1, n° 2, p. 76, 2012.
- [65] G. N. Murshudov *et al.*, "REFMAC5 for the refinement of macromolecular crystal structures", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 67, n° 4, p. 355–367, 2011.
- [66] P. Emsley e K. Cowtan, "Coot: model-building tools for molecular graphics", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 60, n° 12, p. 2126–2132, 2004.
- [67] P. Emsley, B. Lohkamp, W. G. Scott, e K. Cowtan, "Features and development of Coot", *Acta Crystallogr. D Biol. Crystallogr.*, vol. 66, n° 4, p. 486–501, 2010.
- [68] E. F. Pettersen *et al.*, "UCSF Chimera—a visualization system for exploratory research and analysis", *J. Comput. Chem.*, vol. 25, n° 13, p. 1605–1612, 2004.
- [69] A. Wlodawer, W. Minor, Z. Dauter, e M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures", *FEBS J.*, vol. 275, n° 1, p. 1–21, 2008.
- [70] C. Giacovazzo, *Fundamentals of crystallography*, vol. 7. Oxford university press, USA, 2002.
- [71] *Fish Quality: Role of Biological Membranes*. Nordic Council of Ministers, 1995.
- [72] T. Okazaki, R. M. Bell, e Y. A. Hannun, "Sphingomyelin turnover induced by vitamin D3 in HL-60 cells. Role in cell differentiation.", *J. Biol. Chem.*, vol. 264, n° 32, p. 19076–19080, 1989.

- [73] R. N. Kolesnick, "Sphingomyelin and derivatives as cellular signals", *Prog. Lipid Res.*, vol. 30, nº 1, p. 1–38, 1991.
- [74] T. Okazaki, A. Bielawska, N. Domae, R. M. Bell, e Y. A. Hannun, "Characteristics and partial purification of a novel cytosolic, magnesium-independent, neutral sphingomyelinase activated in the early signal transduction of 1 alpha,25-dihydroxyvitamin D3-induced HL-60 cell differentiation", *J. Biol. Chem.*, vol. 269, nº 6, p. 4070–4077, fev. 1994.
- [75] K. A. Dressler, S. Mathias, e R. N. Kolesnick, "Tumor Necrosis Factor-(Alpha) Activates the Sphingomyelin Signal Transduction Pathway in a Cell-Free System", *Science*, vol. 255, nº 5052, p. 1715, 1992.
- [76] K. Venkataraman e A. H. Futerman, "Ceramide as a second messenger: sticky solutions to sticky problems", *Trends Cell Biol.*, vol. 10, nº 10, p. 408–412, 2000.
- [77] A. Gómez-Muñoz, "Ceramide 1-phosphate/ceramide, a switch between life and death", *Biochim. Biophys. Acta BBA-Biomembr.*, vol. 1758, nº 12, p. 2049–2056, 2006.
- [78] G. J. Binford, M. H. Cordes, e M. A. Wells, "Sphingomyelinase D from venoms of *Loxosceles* spiders: evolutionary insights from cDNA sequences and gene structure", *Toxicon*, vol. 45, nº 5, p. 547–560, 2005.
- [79] C. Dias-Lopes *et al.*, "Identification of new sphingomyelinases D in pathogenic fungi and other pathogenic organisms", *PloS One*, vol. 8, nº 11, p. e79240, 2013.
- [80] M. A. M. Corti *et al.*, "Rare human skin infection with *Corynebacterium ulcerans*: transmission by a domestic cat", *Infection*, vol. 40, nº 5, p. 575–578, 2012.
- [81] B. A. Lipsky, A. C. Goldberger, L. S. Tompkins, e J. J. Plorde, "Infections caused by nondiphtheria corynebacteria", *Rev. Infect. Dis.*, vol. 4, nº 6, p. 1220–1235, 1982.
- [82] P. O. de Giuseppe *et al.*, "Structure of a novel class II phospholipase D: catalytic cleft is modified by a disulphide bridge", *Biochem. Biophys. Res. Commun.*, vol. 409, nº 4, p. 622–627, 2011.
- [83] M. A. Coronado *et al.*, "Structural insights into substrate binding of brown spider venom class II phospholipases D", *Curr. Protein Pept. Sci.*, vol. 16, nº 8, p. 768–774, 2015.
- [84] M. T. Murakami, M. F. Fernandes-Pedrosa, D. V. Tambourgi, e R. K. Arni, "Structural basis for metal ion coordination and the catalytic mechanism of sphingomyelinases D", *J. Biol. Chem.*, vol. 280, nº 14, p. 13658–13664, 2005.

- [85] D. M. Lajoie *et al.*, “Variable substrate preference among phospholipase D toxins from sicariid spiders”, *J. Biol. Chem.*, vol. 290, n° 17, p. 10994–11007, 2015.
- [86] J. Monod, “The growth of bacterial cultures”, *Annu. Rev. Microbiol.*, vol. 3, n° 1, p. 371–394, 1949.
- [87] K. D. Kelley, L. Q. Olive, A. Hadziselimovic, e C. R. Sanders, “Look and see if it is time to induce protein expression in *Escherichia coli* cultures”, *Biochemistry (Mosc.)*, vol. 49, n° 26, p. 5405–5407, 2010.
- [88] U. K. Laemmli, “Cleavage of structural proteins during the assembly of the head of bacteriophage T4”, *nature*, vol. 227, n° 5259, p. 680–685, 1970.
- [89] R. G. Taylor, D. C. Walker, e R. R. McInnes, “*E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing.”, *Nucleic Acids Res.*, vol. 21, n° 7, p. 1677, 1993.
- [90] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, e T. Gray, “How to measure and predict the molar absorption coefficient of a protein”, *Protein Sci.*, vol. 4, n° 11, p. 2411–2423, 1995.
- [91] E. Gasteiger *et al.*, *Protein identification and analysis tools on the ExPASy server*. Springer, 2005.
- [92] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, e A. Bairoch, “ExPASy: the proteomics server for in-depth protein knowledge and analysis”, *Nucleic Acids Res.*, vol. 31, n° 13, p. 3784–3788, 2003.
- [93] Y. Zhang, “I-TASSER server for protein 3D structure prediction”, *BMC Bioinformatics*, vol. 9, n° 1, p. 40, 2008.
- [94] D. P. Arfsten, D. J. Schaeffer, e D. C. Mulveny, “The effects of near ultraviolet radiation on the toxic effects of polycyclic aromatic hydrocarbons in animals and plants: a review”, *Ecotoxicol. Environ. Saf.*, vol. 33, n° 1, p. 1–24, 1996.
- [95] R. Bartha, “Biotechnology of petroleum pollutant biodegradation”, *Microb. Ecol.*, vol. 12, n° 1, p. 155–172, 1986.
- [96] N. T. Edwards, “Polycyclic aromatic hydrocarbons (PAH’s) in the terrestrial environment—a review”, *J. Environ. Qual.*, vol. 12, n° 4, p. 427–441, 1983.
- [97] Z. Wang e M. Fingas, “Developments in the analysis of petroleum hydrocarbons in oils, petroleum products and oil-spill-related environmental samples by gas chromatography”, *J. Chromatogr. A*, vol. 774, n° 1, p. 51–78, 1997.
- [98] J. L. Gallego, J. Loredó, J. F. Llamas, F. Vázquez, e J. Sánchez, “Bioremediation of diesel-contaminated soils: evaluation of potential in

- situ techniques by study of bacterial degradation”, *Biodegradation*, vol. 12, nº 5, p. 325–335, 2001.
- [99] S. M. Bamforth e I. Singleton, “Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions”, *J. Chem. Technol. Biotechnol.*, vol. 80, nº 7, p. 723–736, 2005.
- [100] T. Shen, Y. Pi, M. Bao, N. Xu, Y. Li, e J. Lu, “Biodegradation of different petroleum hydrocarbons by free and immobilized microbial consortia”, *Environ. Sci. Process. Impacts*, vol. 17, nº 12, p. 2022–2033, 2015.
- [101] A. K. Haritash e C. P. Kaushik, “Biodegradation aspects of polycyclic aromatic hydrocarbons (PAHs): a review”, *J. Hazard. Mater.*, vol. 169, nº 1, p. 1–15, 2009.
- [102] C. E. Cerniglia, “Fungal metabolism of polycyclic aromatic hydrocarbons: past, present and future applications in bioremediation”, *J. Ind. Microbiol. Biotechnol.*, vol. 19, nº 5, p. 324–333, 1997.
- [103] S. K. Samanta, O. V. Singh, e R. K. Jain, “Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation”, *TRENDS Biotechnol.*, vol. 20, nº 6, p. 243–248, 2002.
- [104] S. M. Bamforth e I. Singleton, “Bioremediation of polycyclic aromatic hydrocarbons: current knowledge and future directions”, *J. Chem. Technol. Biotechnol.*, vol. 80, nº 7, p. 723–736, 2005.
- [105] M. Sota *et al.*, “Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase”, *J. Bacteriol.*, vol. 188, nº 11, p. 4057–4067, 2006.
- [106] S. L. Guimarães, “NahK/NahL: Estrutura, função e mecanismo de ação de um complexo macromolecular”, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.
- [107] Z. Otwinowski e W. Minor, “Processing of X-ray diffraction data collected in oscillation mode”, *Methods Enzymol.*, vol. 276, p. 307–326, 1997.
- [108] G. M. Sheldrick, “A short history of SHELX”, *Acta Crystallogr. A*, vol. 64, nº 1, p. 112–122, 2008.
- [109] E.-M. De Villiers, C. Fauquet, T. R. Broker, H.-U. Bernard, e H. zur Hausen, “Classification of papillomaviruses”, *Virology*, vol. 324, nº 1, p. 17–27, 2004.
- [110] H. Trottier e A. N. Burchell, “Epidemiology of mucosal human papillomavirus infection and associated diseases”, *Public Health Genomics*, vol. 12, nº 5–6, p. 291–307, 2009.
- [111] D. Forman *et al.*, “Global burden of human papillomavirus and related diseases”, *Vaccine*, vol. 30, p. F12–F23, 2012.

- [112] K. Münger e P. M. Howley, "Human papillomavirus immortalization and transformation functions", *Virus Res.*, vol. 89, n° 2, p. 213–228, 2002.
- [113] B. A. Spalholz, Y.-C. Yang, e P. M. Howley, "Transactivation of a bovine papilloma virus transcriptional regulatory element by the E2 gene product", *Cell*, vol. 42, n° 1, p. 183–191, 1985.
- [114] R. S. Hegde, S. R. Grossman, L. A. Laimins, e P. B. Sigler, "Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target", *Nature*, vol. 359, n° 6395, p. 505–512, 1992.
- [115] R. S. Hegde, A.-F. Wang, S.-S. Kim, e M. Schapira, "Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein", *J. Mol. Biol.*, vol. 276, n° 4, p. 797–808, 1998.
- [116] S. Veeraraghavan, C. C. Mello, E. J. Androphy, e J. D. Baleja, "Structural Correlates for Enhanced Stability in the E2 DNA-Binding Domain from Bovine Papillomavirus†", *Biochemistry (Mosc.)*, vol. 38, n° 49, p. 16115–16124, 1999.
- [117] J. Barbosa *et al.*, "Crystallization and X-Ray Data Analysis of the Extended Dna-Binding Domain of the E2 Bovine Papillomavirus Type 1 Protein", *Protein Pept. Lett.*, vol. 8, n° 4, p. 323–326, 2001.
- [118] B. W. Matthews, "Solvent content of protein crystals", *J. Mol. Biol.*, vol. 33, n° 2, p. 491–497, 1968.
- [119] K. A. Kantardjieff e B. Rupp, "Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals", *Protein Sci.*, vol. 12, n° 9, p. 1865–1871, 2003.
- [120] M. P. Jacobson, R. A. Friesner, Z. Xiang, e B. Honig, "On the role of the crystal environment in determining protein side-chain conformations", *J. Mol. Biol.*, vol. 320, n° 3, p. 597–608, 2002.
- [121] C. Interest, "A Brief Guide to the Twenty Common Amino Acids", *Compound Interest*, 16-set-2014.